



IJCSIS Vol. 16 No. 5, May 2018
ISSN 1947-5500

International Journal of Computer Science & Information Security

© IJCSIS PUBLICATION 2018
Pennsylvania, USA

Indexed and technically co-sponsored by :



AUTHOR SERIES



Indexing Service

IJCSIS has been indexed by several world class databases, for more information, please access the following links:

Global Impact Factor

<http://globalimpactfactor.com/>

Google Scholar

<http://scholar.google.com/>

CrossRef

<http://www.crossref.org/>

Microsoft Academic Search

<http://academic.research.microsoft.com/>

IndexCopernicus

<http://journals.indexcopernicus.com/>

IET Inspec

<http://www.theiet.org/resources/inspec/>

EBSCO

<http://www.ebscohost.com/>

JournalSeek

<http://journalseek.net>

Ulrich

<http://ulrichsweb.serialssolutions.com/>

WordCat

<http://www.worldcat.org>

Academic Journals Database

<http://www.journaldatabase.org/>

Stanford University Libraries

<http://searchworks.stanford.edu/>

Harvard Library

<http://discovery.lib.harvard.edu/?itemid=|library/m/aleph|012618581>

UniSA Library

<http://www.library.unisa.edu.au/>

ProQuest

<http://www.proquest.co.uk>

Zeitschriftendatenbank (ZDB)
<http://dispatch.opac.d-nb.de/>

IJCSIS

ISSN (online): 1947-5500

Please consider to contribute to and/or forward to the appropriate groups the following opportunity to submit and publish original scientific results.

CALL FOR PAPERS

International Journal of Computer Science and Information Security (IJCSIS) January-December 2018 Issues

The topics suggested by this issue can be discussed in term of concepts, surveys, state of the art, research, standards, implementations, running experiments, applications, and industrial case studies. Authors are invited to submit complete unpublished papers, which are not under review in any other conference or journal in the following, but not limited to, topic areas.

See authors guide for manuscript preparation and submission guidelines.

Indexed by Google Scholar, DBLP, CiteSeerX, Directory for Open Access Journal (DOAJ), Bielefeld Academic Search Engine (BASE), SCIRUS, Scopus Database, Cornell University Library, ScientificCommons, ProQuest, EBSCO and more.

Deadline: see web site

Notification: see web site

Revision: see web site

Publication: see web site

Context-aware systems
Networking technologies
Security in network, systems, and applications
Evolutionary computation
Industrial systems
Evolutionary computation
Autonomic and autonomous systems
Bio-technologies
Knowledge data systems
Mobile and distance education
Intelligent techniques, logics and systems
Knowledge processing
Information technologies
Internet and web technologies, IoT
Digital information processing
Cognitive science and knowledge

Agent-based systems
Mobility and multimedia systems
Systems performance
Networking and telecommunications
Software development and deployment
Knowledge virtualization
Systems and networks on the chip
Knowledge for global defense
Information Systems [IS]
IPv6 Today - Technology and deployment
Modeling
Software Engineering
Optimization
Complexity
Natural Language Processing
Speech Synthesis
Data Mining

For more topics, please see web site <https://sites.google.com/site/ijcsis/>

arXiv.org Google scholar

SCIRUS
search engine for science

ScientificCommons

Scribd

docstoc
find and share professional documents

BASE
Bielefeld Academic Search Engine

CiteSeer^x beta

dblp.uni-trier.de
Computer Science
Bibliography

DOAJ
DIRECTORY OF
OPEN ACCESS
JOURNALS

EBSCO
HOST

ProQuest

For more information, please visit the journal website (<https://sites.google.com/site/ijcsis/>)

Editorial Message from Editorial Board

*It is our immense pleasure to present the **May 2018 issue** (Volume 16 Number 5) of the **International Journal of Computer Science and Information Security (IJCSIS)**. High quality research, survey & review articles are proposed from experts in the field, promoting insight and understanding of the state of the art, and trends in computer science and digital technologies. It especially provides a platform for high-caliber academics, practitioners and PhD/Doctoral graduates to publish completed work and latest research outcomes. According to Google Scholar, up to now papers published in IJCSIS have been cited over 10751 times and this journal is experiencing steady and healthy growth. Google statistics shows that IJCSIS has established the first step to be an international and prestigious journal in the field of Computer Science and Information Security. There have been many improvements to the processing of papers; we have also witnessed a significant growth in interest through a higher number of submissions as well as through the breadth and quality of those submissions. IJCSIS is indexed in major academic/scientific databases and important repositories, such as: Google Scholar, Thomson Reuters, ArXiv, CiteSeerX, Cornell's University Library, Ei Compendex, ISI Scopus, DBLP, DOAJ, ProQuest, ResearchGate, LinkedIn, Academia.edu and EBSCO among others.*

A great journal has a dedicated editorial team of editors and reviewers. On behalf of IJCSIS community and the sponsors, we congratulate the authors and thank the reviewers & editors for their outstanding efforts to meticulously review and recommend high quality papers for publication. In particular, we would like to thank the international academia and researchers for continued support by citing or reading papers published in IJCSIS. Without their sustained and unselfish commitments, IJCSIS would not have achieved its current premier status, making sure we deliver high-quality content to our readers in a timely fashion.

"We support researchers to succeed by providing high visibility & impact value, prestige and excellence in research publication." We would like to thank you, the authors and readers, the content providers and consumers, who have made this journal the best possible.

For further questions or other suggestions please do not hesitate to contact us at ijcsiseditor@gmail.com.

*A complete list of journals can be found at:
<http://sites.google.com/site/ijcsis/>*

IJCSIS Vol. 16, No. 5, May 2018 Edition

ISSN 1947-5500 © IJCSIS, USA.

Journal Indexed by (among others):



Open Access This Journal is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source.



Bibliographic Information

ISSN: 1947-5500

Monthly publication (Regular Special Issues)
Commenced Publication since May 2009

Editorial / Paper Submissions:

IJCSIS Managing Editor

ijcsiseditor@gmail.com

Pennsylvania, USA

Tel: +1 412 390 5159

IJCSIS EDITORIAL BOARD

| IJCSIS Editorial Board | IJCSIS Guest Editors / Associate Editors |
|--|---|
| Dr. Shimon K. Modi [Profile] Director of Research BSPA Labs, Purdue University, USA | Dr Riktesh Srivastava [Profile] Associate Professor, Information Systems, Skyline University College, Sharjah, PO 1797, UAE |
| Professor Ying Yang, PhD. [Profile] Computer Science Department, Yale University, USA | Dr. Jianguo Ding [Profile] Norwegian University of Science and Technology (NTNU), Norway |
| Professor Hamid Reza Naji, PhD. [Profile] Department of Computer Enigneering, Shahid Beheshti University, Tehran, Iran | Dr. Naseer Alquraishi [Profile] University of Wasit, Iraq |
| Professor Yong Li, PhD. [Profile] School of Electronic and Information Engineering, Beijing Jiaotong University, P. R. China | Dr. Kai Cong [Profile] Intel Corporation, & Computer Science Department, Portland State University, USA |
| Professor Mokhtar Beldjehem, PhD. [Profile] Sainte-Anne University, Halifax, NS, Canada | Dr. Omar A. Alzubi [Profile] Al-Balqa Applied University (BAU), Jordan |
| Professor Yousef Farhaoui, PhD. Department of Computer Science, Moulay Ismail University, Morocco | Dr. Jorge A. Ruiz-Vanoye [Profile] Universidad Autónoma del Estado de Morelos, Mexico |
| Dr. Alex Pappachen James [Profile] Queensland Micro-nanotechnology center, Griffith University, Australia | Prof. Ning Xu, Wuhan University of Technology, China |
| Professor Sanjay Jasola [Profile] Gautam Buddha University | Dr . Bilal Alatas [Profile] Department of Software Engineering, Firat University, Turkey |
| Dr. Siddhivinayak Kulkarni [Profile] University of Ballarat, Ballarat, Victoria, Australia | Dr. Ioannis V. Koskosas, University of Western Macedonia, Greece |
| Dr. Reza Ebrahimi Atani [Profile] University of Guilan, Iran | Dr Venu Kuthadi [Profile] University of Johannesburg, Johannesburg, RSA |
| Dr. Dong Zhang [Profile] University of Central Florida, USA | Dr. Zhihan Iv [Profile] Chinese Academy of Science, China |
| Dr. Vahid Esmaeelzadeh [Profile] Iran University of Science and Technology | Prof. Ghulam Qasim [Profile] University of Engineering and Technology, Peshawar, Pakistan |
| Dr. Jiliang Zhang [Profile] Northeastern University, China | Prof. Dr. Maqbool Uddin Shaikh [Profile] Preston University, Islamabad, Pakistan |
| Dr. Jacek M. Czerniak [Profile] Casimir the Great University in Bydgoszcz, Poland | Dr. Musa Peker [Profile] Faculty of Technology, Mugla Sitki Kocman University, Turkey |
| Dr. Binh P. Nguyen [Profile] National University of Singapore | Dr. Wencan Luo [Profile] University of Pittsburgh, US |
| Professor Seifeidne Kadry [Profile] American University of the Middle East, Kuwait | Dr. Ijaz Ali Shoukat [Profile] King Saud University, Saudi Arabia |
| Dr. Riccardo Colella [Profile] University of Salento, Italy | Dr. Yilun Shang [Profile] Tongji University, Shanghai, China |
| Dr. Sedat Akleylek [Profile] Ondokuz Mayıs University, Turkey | Dr. Sachin Kumar [Profile] Indian Institute of Technology (IIT) Roorkee |

| | |
|---|--|
| Dr Basit Shahzad [Profile] King Saud University, Riyadh - Saudi Arabia | Dr. Mohd. Muntjir [Profile] Taif University Kingdom of Saudi Arabia |
| Dr. Sherzod Turaev [Profile] International Islamic University Malaysia | Dr. Bohui Wang [Profile] School of Aerospace Science and Technology, Xidian University, P. R. China |
| Dr. Kelvin LO M. F. [Profile] The Hong Kong Polytechnic University, Hong Kong | Dr. Man Fung LO [Profile] The Hong Kong Polytechnic University |

TABLE OF CONTENTS

1. PaperID 30041811: Detection of Seam Carving in Uncompressed Images using eXtreme Gradient Boosting (pp. 1-5)

Odysseas B. Efremides, School of ICT and Web Media, Bahrain Polytechnic, PO Box 33349, Isa Town, Kingdom of Bahrain

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

2. PaperID 30041804: Simulation for Data Security Improvement in Exploited Metarouter (pp. 6-15)

*Kristono, Department of Informatics, Universitas Islam Indonesia
Imam Riadi, Department of Information System, Universitas Ahmad Dahlan*

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

3. PaperID 30041805: Optimization Algorithm to Control Interference-based Topology Control for Delay-Constrained Mobile Ad-hoc Networks (pp. 16-22)

*Selvakumar (1), S. Selvakumar (2)
(1) Research Scholar, Department of Computer Science and Engineering, SRM Institute of Science and Technology,
(2) Assistant Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology*

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

4. PaperID 30041809: Intelligent Intrusion Detection System Based on MLP, RBF and SVM Classification Algorithms: A Comparative Study (pp. 23-30)

Ismail M. Keshta, Department of Computer and Information Technology, Dammam Community College (DCC), King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia.

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

5. PaperID 30041810: An Enhanced Approach of Clustering Protocol to Minimize Energy Holes in Wireless sensor networks (WSN) (pp. 31-37)

*Chinmaya Kumar Nayak & Satyabrata Das
Department of CSE & IT, Veer Surendra Sai University and Technology (VSSUT), Burla, Odisha, India*

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

6. PaperID 30041813: Enhanced Hybrid Blowfish and ECC Encryption to Secure Cloud Data Access and Storage Policies (pp. 38-44)

*(1) Jobandeep Kaur, (2) Dr. Vishal Bharti, (3) Mr. Shamandeep Singh
(1) M. Tech (Scholar), (2) Head of Department, (3) Assistant Professor
Department of Computer Science and Engineering, Chandigarh, Gharuan, Punjab (INDIA)*

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

7. PaperID 30041817: The Challenges of E-Governance Implementation in Nigerian Aviation Industry (pp. 45-52)

Eleberi Ebele Leticia, Department of Computer Science, Imo State University, PMB 2000, Owerri Nigeria
Ekwowunne Emmanuel, Department of Computer Science, Imo State University, PMB 2000, Owerri Nigeria

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

8. PaperID 30041818: Ensemble Feature Selection (EFS) and Ensemble Convolutional Neural Network with Vote Classification (ECNN-VC) Approach for Detection of Snore Sounds in Obstructive Sleep Apnea Diagnosis (pp. 53-63)

E. Jeslin Renjith, Research Scholar, Bharathiar University, Coimbatore, India
Dr. Christy A., Research Supervisor, Bharathiar University, Coimbatore, India

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

9. PaperID 30041819: Leaf Recognition System for Plant Identification and Classification Based On Leaf Features Fusion & Zernike Moment (pp. 64-73)

Jignesh B. Jethva & Prof. Dr. C. R. Singla
Electronics and Communication Department, Madhav University, Abu Road, Pindwara, Dist. Sirohi, Bharja, Rajasthan, India

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

10. PaperID 30041823: A Hybrid Approach for Classification of MRI Brain Tumors Using Genetic Algorithm, K-Nearest Neighbor and Probabilistic Neural Network (pp. 74-85)

Raghad Majeed Azawi, Department of Computer Science, College of Science, University of Diyala, Diyala, Iraq
Dhahir Abdulhade Abdulah, Department of Computer Sciences, College of Science, University of Diyala, Diyala, Iraq
Jamal Mustafa Abbas, Department of Computer Science, College of Science, University of Diyala, Diyala, Iraq
Ibrahim Tareq Ibrahim, College of Medicine, University of Diyala, Diyala, Iraq

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

11. PaperID 30041826: Information Retrieval thru Cellular Devices (pp. 86-93)

Muhammad Zunnurain Hussain, Sr. Lecturer CS & IT, Bahria University, Lahore Campus, Pakistan.
Tahir Iqbal, Sr. Assistant Professor, Bahria University, Lahore Campus, Pakistan.
Muhammad Zulkifl Hasan, Sr. Lecturer CS & IT, Lahore Garrison University Lahore, Pakistan.
Adeel. M, Student 6th Sem, LGU Lahore, Pakistan.

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

12. PaperID 30041827: A Security Overview of Wireless Sensor Network (pp. 94-100)

Muhammad Zulkifl Hasan (1), Zaka Ullah (2), Muhammad Zunnurain Hussain (3)*
[1, 2] Department of Computer Science, Lahore Garrison University, Lahore, Pakistan,
[3] Bahria University Lahore Campus, Lahore Pakistan

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

13. PaperID 30041832: An Exposition of Performance Comparison of Graphic Processing Unit Virtualization Techniques (pp. 101-104)

Asif Farooq, Lecturer, Department of Computer Science and Information Technology, University of Sargodha, Lahore Campus, Pakistan
Tahir Iqbal, Sr. Assistant Professor, Department of Computer Science and Information Technology, Bahria University, Lahore Campus, Pakistan

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

14. PaperID 30041833: User Feedback based Ranking for Yahoo Answers (pp. 105-117)

Muhammad Awais Shaheen, Department of Computer Sciences, International Islamic University, Islamabad, Pakistan.

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

15. PaperID 30041840: ACSR Overhead Transmission lines Modeling Considering Sag-Tension Estimation in Cold Area with different Wind Speeds (pp. 118-121)

Muhammad Zulqarnain Abbasi & Hamza Umar Afridi
IQRA National University, Pakistan

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

16. PaperID 30041841: Comparison of Machine Learning Techniques for Identification of Disease (pp. 122-130)

Shaveta Malik (1), Tapas Kumar (2), A.K. Sahoo (3)*
(1, 2) School of Computer Science, Lingaya's University, Faridabad, Haryana.
(3) Scientist, Biodiversity Lab, Reverie Ecology and Fisheries Division, Barrackpore, Kolkata.*

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

17. PaperID 30041843: SiDe Enabled Reliable Replica Optimization (pp. 131-139)

Prof. Anjali Chandavale
Yash Palwe, student of MITCOE, Pune, India.
KartikHedau, student of MITCOE, Pune, India.
Akshay Ghorpade, student of MITCOE, Pune, India.
Ayushi Dantulwar, student of MITCOE, Pune, India.

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

18. PaperID 30041850: Wireless Automated Soil Monitoring System (pp. 140-145)

Sadia Afrin, Dept. of CSE, Bangladesh University of Engineering & Technology, (BUET) Dhaka, Bangladesh
Shampa Buraira, .NET Developer, Emerald Techno Limited, Dhaka, Bangladesh
Nasir Umme Salma, Lecturer, Dept. of CSE, Bangladesh University, Dhaka, Bangladesh

Full Text: [PDF](#) [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

19. PaperID 30041856: Cloudlet based Healthcare and Medical Knowledge Extraction System for Medical Big Data (pp. 146-152)

Dr. Anjali A. Chandavale, Department of Information Technology, MIT College of Engineering, Pune, India
Anuja J. Gade, Department of Information Technology, MIT College of Engineering, Pune, India
Anuja J. Gade, Department of Information Technology, MIT College of Engineering, Pune, India
Dr. Anjali A. Chandavale, Department of Information Technology, MIT College of Engineering, Pune, India

Full Text: [PDF](#) [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

20. PaperID 30041868: FiDooP: Parallel Mining of Frequent Itemsets Using MapReduce (pp. 153-157)

Dr G. Krishna Kishore, Computer Science and Engineering, V. R. Siddhartha Engineering College, Vijayawada, Andhra Pradesh, India
Suresh Babu Dasari, Computer Science and Engineering, V. R. Siddhartha Engineering College, Vijayawada, Andhra Pradesh, India
S. Ravi Kishan, Computer Science & Engineering, V.R. Siddhartha Engineering College, Vijayawada, Andhra Pradesh, India

Full Text: [PDF](#) [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

21. PaperID 30041869: Pleomorphism in Cervical Nucleus: A Review (pp. 158-162)

Jing Rui TANG, Engineering Technology Department, Faculty of Technical and Vocational, Sultan Idris Education University, 35900 Tanjong Malim, Perak, Malaysia

Full Text: [PDF](#) [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

22. PaperID 30041871: Comparative Study of Diverse API Perspective of Spatial Data (pp. 163-168)

Zainab Tariq (1) & Umesh Chandra (2)
(1) Research Associate, Computer Science, Glocal University, Saharanpur
(2) Assistant Professor, Computer Science, Banda University of Agriculture and Technology, Banda

Full Text: [PDF](#) [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

23. PaperID 30041872: Survey of Feature Selection / Extraction Methods used in Biomedical Imaging (pp. 169-177)

Misbah Ziaye, CSIT Department, MUST University (Mirpur, AJK)
Samina Khalid, CSIT Department, MUST University (Mirpur, AJK)
Yasir Mehmood, CSIT Department, MUST University (Mirpur, AJK)

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

24. PaperID 30041873: A Novel for Finding Missing Piece in the Borderless Square Jigsaw Puzzle (pp. 178-194)

Sendi Novianto & Luo Fei

School of Automation, South China University of Technology, Guangzhou – Guangdong, China

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

25. PaperID 30041876: Identification and Recognition of Snack Foods from Cluttered Scene for Dietary Assessment (pp. 195-200)

Amatul Bushra Akhi, Farzana Akter & Tania Khatun

Department of Computer Science and Engineering, Daffodil International University, Dhaka

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

26. PaperID 30041879: Evidence Gathering and Identification of LINE Messenger on Android Device (pp. 201-205)

Imam Riadi, Department of Information Systems, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

Abdul Fadlil, Department of Electronic Engineering, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

Ammar Fauzan, Master of Informatics Engineering, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

27. PaperID 30041887: An Overview of Biometric Template Security Techniques (pp. 206-210)

Lilian Imali Madivoli,

Institute of Computer Science and Information Technology, Jomo Kenyatta University of Agriculture and Technology (JKUAT), Nairobi, Kenya

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

28. PaperID 30041892: Design and Implementation of a Language Assistant for English – Arabic Texts (pp. 211-221)

Akeem Alabi (1), Oluwatobi Ayilara (2), Kudirat Jimoh (3)

(1) College of Technology, The Polytechnic. Ile-Ife, Osun State, Nigeria

(2) College of Technology, Oduduwa University, Ipetumodu, Osun State, Nigeria

(3) Natural and Applied Sciences, The Federal Polytechnic, Ede, Osun State, Nigeria

Full Text: PDF [[Academia.edu](#) | [Scopus](#) | [Scribd](#) | [Archive](#) | [ProQuest](#)]

29. PaperID 30041894: Mammographic Image Enhancement using Digital Image Processing Technique (pp. 222-226)

Ardymulya Iswardani, Information System Department, Stmik Duta Bangsa, Surakarta, Indonesia

Wahyu Hidayat, Informatics Department, Stmik Duta Bangsa, SURAKARTA, Indonesia

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

30. PaperID 30041895: Improved K-mean Clustering Algorithm for Prediction Analysis using Classification Technique in Data Mining (pp. 227-239)

Swati Vinodani, Galgotia College of Engineering
Aatif Jamshed, Assistant professor, Galgotia college of engineering
Dr. Pramod Kumar, Director, Tula's Institute of Technology

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

31. PaperID 30041896: Lecturers' Attitudes towards Integration of E-Learning in Higher Education. Case Study: University of Tetovo (pp. 240-249)

Gazmend Xhaferi, University of Tetovo, Tetovo, Macedonia
Rovena Bahiti, University of Tirana, Tirana, Albania

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

32. PaperID 30041897: Simulation Based Workflow Scheduling for Scientific Application (pp. 250-256)

Priya Garg & Dr. Anju Bala
Computer and Science Engineering Department, Thapar University, Patiala 147004, India

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

33. PaperID 310318104: Speed Control of Three Phase Induction Motor Using Neural Network (pp. 257-272)

Eng. Nada Mohammed Abd Elmonem Mahmoud - Faculty of Engineering-Computers Engineering & Control Systems Department- Mansoura University. Demonstrator of Computer at Nile Higher Institute for Commercial Science and Computer Technology. (A.R.E).
Dr. Mohamed Moawad Abdu- Faculty of Engineering-Computers Engineering & Control Systems Department- Mansoura University.
Assoc.Prof. Mohammed Sherif Moustafa - Faculty of Engineering-Computers Engineering & Control Systems Department- Mansoura University.
Assoc.Prof. Sabry Fouad Saraya- Faculty of Engineering-Computers Engineering & Control Systems Department- Mansoura University.

Full Text: PDF [Academia.edu | Scopus | Scribd | Archive | ProQuest]

Detection of Seam Carving in Uncompressed Images using eXtreme Gradient Boosting

Odysseas B. Efremides

School of ICT and Web Media, Bahrain Polytechnic
PO Box 33349, Isa Town, Kingdom of Bahrain
odyssefs.effraimidis@polytechnic.bh

Abstract—A digital image forensic approach to detect whether an image has been seam carved or not is investigated herein. Seam carving is a content-aware image retargeting technique which preserves the semantically important content of an image while resizing it. The same technique, however, can be used for malicious tampering of an image. 18 energy, seam, and noise related features defined by Ryu [1] are produced using Sobel's [2] gradient filter and Rubinstein's [3] forward energy criterion enhanced with image gradients. An extreme gradient boosting classifier [4] is trained to make the final decision. Experimental results show that the proposed approach improves the detection accuracy from 5 to 10% for seam carved images with different scaling ratios when compared with other state-of-the-art methods.

Index Terms—digital image forensics, seam carving, extreme gradient boosting, content-aware image resizing

I. INTRODUCTION

Digital image forensics aims at validating the authenticity of images. To this goal, a number of passive-blind techniques have been developed during the last decade. *Passive* means that these techniques require no access to the image capturing device while *blind* means that they do not need to know anything about the original or any other intermediate image produced during the process [5]. All these methods assume that manipulating an image creates artifacts in the resulting image by disturbing the statistical properties of the original one. Thus, by examining these statistical properties artifacts indicating manipulation might be found. To detect these artifacts a number of different detectors have been proposed. Most of those detectors [6], [7] assume that the entire image is altered during the retargeting resampling process and, thus, they fail to detect artifacts introduced by methods like the seam carving resizing technique.

Seam carving proposed by Avidan and Shamir [8] in 2007 as a content aware resizing technique. In order to resize, change the aspect ratio or intentionally carved out some parts of a digital image while preserving any important content in it they delete low energy pixels which might be considered as unnoticeable or less important. This process creates artifacts into the final image that can be used for forensic purposes.

The rest of the paper is organized as follows. In Section II, the seam carving technique is presented followed by a brief description of detectors found in literature. 3 state-of-the-art approaches, used for comparison with the proposed approach, are also described in detail in this section. In Section III, the

proposed approach is presented while its experimental results are reported in Section IV. Conclusive remarks can be found in Section V.

II. BACKGROUND

A. Seam Carving Process Overview

Seam carving [8] proposed as a novel content-aware image retargeting method. That means that semantically important part (interest parts) in an image are not affected by its resizing process (or affected the least possible). A seam is defined as an 8-connected path of low energy pixels crossing the image from top to bottom (vertical seam), or from left to right (horizontal seam). A vertical seam for an $n \times m$ image I is defined by Eq. 1 as:

$$s_i^v = (x(i), i)_{i=1}^n, s.t. \forall i, |x(i) - x(i-1)| \leq 1 \quad (1)$$

while a horizontal seam can be defined similarly. By successively removing unnoticeable seams, seams bearing minimum energy, the important image content can be preserved during the resizing process.

In order for the seam with the minimum energy to be found a function/operator measuring image energy at each pixel is defined by Eq. 2

$$e(I) = \left| \frac{\partial}{\partial x} I \right| + \left| \frac{\partial}{\partial y} I \right| \quad (2)$$

Given this energy function e the cost of each vertical seam s is defined by Eq. 3:

$$E(s) = E(I_s) = \sum_{i=1}^n e(I(s_i)) \quad (3)$$

The optimal seam s^* minimizes the cost:

$$s^* = \min_s E(s) \quad (4)$$

and can be found by building a cumulative minimum energy matrix M for all possible connected seams for each entry (i, j) using dynamic programming.

At the end of the process the minimum value in the last row of M indicates the endpoint of the minimal connected vertical seam. Hence, in a second step, a backtracking process starts from that minimal entry in M towards the top of the matrix in order for the optimal seam path to be found and removed. Figure 1 shows 11 seams found under a 3% vertical reduction of the image size of size 384×512 pixels.



Fig. 1. Original Image (UCID) (left). 3% Seams to be carved (right)

B. Passive-Blind Seam Carving Detectors

Very few forensic passive-blind detectors dealing with the seam carving resizing process have been proposed. Sarkar et al. [9] used Markov features of DCT coefficients to detect seam carving in JPEG compressed images while Fillion et al. [10] extracted a number of statistical features describing carved images. Based on this work a different number and types of statistical features were exploited [11], [12], [13]. Nine predictive patches, a patch index and a reference pattern proposed by Wei et al. [14] as an alternative approach.

Three state-of-the-art approaches are chosen to be compared with the proposed approach.

Ryu et al. [1] proposed a three-category set of features characterizing an image. Based on the fact that a seam carved image shows a higher energy distribution and that the method removes row or column pixels they measure the pixel energy by four statistical features. More specifically, the row and column average energy, the average energy of the entire image and the difference between column and row energy are calculated.

The second group of measures deals with the seams themselves based on the idea that it is highly probable the energy of the remaining seams in a seam carved image to be higher than that of the original non-carved one. They construct the cumulative minimum energy matrix M for all possible seams (vertical and horizontal) and they compute five statistics values (min, max, mean, standard deviation, and the difference between maximum and minimum values) for both directions. The matrix M is calculated using the backward energy formula described by Eq. 5. That way 10 additional features are produced.

$$M(i, j) = e(i, j) + \min(M(i-1, j-1), M(i-1, j), M(i-1, j+1)) \quad (5)$$

Finally, from the observation that the noise level of a seam carved image is to be affected from the removal of its flat regions they extract the last four statistical features regarding noise. To isolate the noise, N , they filtered an image, I , by a Wiener filter, F , having a window size of 5×5 and then they

compute N using Eq. 6

$$N = I - F(I) \quad (6)$$

The mean, standard deviation, skewness, and kurtosis of the noise are the four statistics computed.

A total of 18 features are extracted and they are presented in Table I.

TABLE I
FEATURE DESCRIPTION

$$\begin{aligned} colEnergy_{aver} &= \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n \left| \frac{\partial}{\partial x} I(i, j) \right| \\ rowEnergy_{aver} &= \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n \left| \frac{\partial}{\partial y} I(i, j) \right| \\ energy_{aver} &= \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n \left(\left| \frac{\partial}{\partial x} I(i, j) \right| + \left| \frac{\partial}{\partial y} I(i, j) \right| \right) \\ diffEnergy_{aver} &= \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n \left(\left| \frac{\partial}{\partial x} I(i, j) \right| - \left| \frac{\partial}{\partial y} I(i, j) \right| \right) \\ max_{vert} &= \max_{i=1}^m M(i, n) \\ min_{vert} &= \min_{i=1}^m M(i, n) \\ mean_{vert} &= \frac{1}{m} \sum_{i=1}^m M(i, n) \\ stdev_{vert} &= \sqrt{\frac{1}{m} \sum_{i=1}^m (mean_{vert} - M(i, n))^2} \\ diff_{vert} &= max_{vert} - min_{vert} \\ max_{horiz} &= \max_{i=1}^n M(m, i) \\ min_{horiz} &= \min_{i=1}^n M(m, i) \\ mean_{horiz} &= \frac{1}{n} \sum_{i=1}^n M(m, i) \\ stdev_{horiz} &= \sqrt{\frac{1}{n} \sum_{i=1}^n (mean_{horiz} - M(m, i))^2} \\ diff_{horiz} &= max_{horiz} - min_{horiz} \\ mean_{noise} &= \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n N(i, j) \\ stdev_{noise} &= \sqrt{\frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n (N(i, j) - N_{mean})^2} \\ skewness_{noise} &= \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n \left(\frac{N(i, j) - N_{mean}}{N_{std}} \right)^3 \\ kurtosis_{noise} &= \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n \left(\frac{N(i, j) - N_{mean}}{N_{std}} \right)^4 \end{aligned}$$

In order to test their approach the UCID [15] image dataset is used. They carved the images from 10% to 50% in steps of 10% and computed the 18 features described above. Afterwards, the feature vectors from all seam-carved and non-carved images were used to train and test a Support Vector Machine (SVM) classifier. As it is described in their paper, the optimization process led them to the following hyperparameter

setting: Radial Basis Function (RBF) kernel of $r = 0.125$ and $C = 3$.

Yin et al. [16] build upon Ryu et.al [1] method. Their process calculates the same 18 features shown in Table I but not from the original image. Instead it uses an image resulted from a local binary pattern (LBP) detector. That way the method incorporate local texture changes to the feature vector. Moreover, the concept of half-seams is defined and 6 new half-seam features are produced to capture energy changes in half images. The construction of the cumulative energy matrix M is done using the backward energy criterion [8] and the final number of features calculated reaches 24. They use the same UCID [15] image set for their experiments and create two different subsets. In the first set the images are carved based on small percentages (from 3% to 21% in steps of 3%) while in the second large reduction percentages (from 10% to 50% in steps of 10%) are used. An SVM is trained as a classifier using an RBF kernel and 3-fold validation but the rest of the parameters are not defined in their paper.

A local derivative pattern (LDP) based forensic approach proposed by Ye et al. [17]. They use the same 24 features described before ([1], [16]) but they extracted them from four different images. Four LDP encoders are applied to encode the original under investigation image producing the four LDP image form which the features are extracted and the process results to 96 different features. For their calculation the backward energy criterion is used and an SVM classifier is utilized. Concerning the setting they state that they use a linear kernel leaving all the other parameters to their default values. UCID [15] image set is used for their experiments and they work with one set of images carved using percentages from 10% to 50% in steps of 10% plus an extra set carved by 5%.

All three groups report their results mainly as a binary classification problem between the non-carved and the carved version of their images (e.g., non-carved vs. 9% or non-carved vs. 50% vertical carving) for each percentage separately. Also they report results on a mixed set of image.

III. THE METHOD

As described in Section II the three state-of-the-art methods used for comparison reasons herein are built one upon the other. The key points in all of them are the number of features used, the criterion used to build the cumulative energy matrix M and the classifier used. The same basic set of features is extended or it is produced from different image representations (LBP, LDP). This increases the number of features used for the classification from 18 to 96. Moreover, in order to capture the image characteristics related to seams these detectors produce their cumulative energy (cost) matrix using the backward criterion as proposed by Avidan et al. [8]. Finally, after extracting their features they utilize an SVM classifier for making their binary classification decision: non-carved, carved.

The pipeline proposed in this work is depicted in Figure 2. The image features used are those proposed by Ryu [1]. That

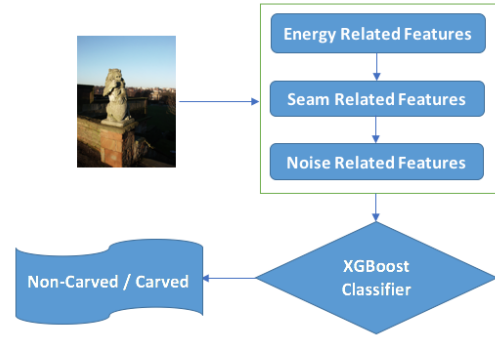


Fig. 2. Classification Pipeline (image form UCID dataset)

means that only 18 such features are used as they are described in Table I. All seam related features are extracted using the forward energy criterion proposed by Rubinstein et al. [3].

In an attempt to eliminate the artifacts in the retargeted image produced by the original algorithm Rubinstein et al. [3] proposed the *forward energy* criterion for the selection of the optimal seam. The idea is that removing a seam brings together previously non adjacent pixels. These pixels, now neighbors, form new edges which add a new amount of energy into the image. Thus, the algorithm looks forward at the the image resulted after removing a seam and chooses to remove the seam whose removal adds the minimum amount of new energy into the retargeted image.

Their cumulative cost matrix M is calculated using dynamic programming. For vertical seams, each cost $M(i, j)$ is updated using the rule given by Eq. 7-10

$$M(i, j) = P(i, j) + \min \begin{cases} M(i-1, j-1) + C_L(i, j) \\ M(i-1, j) + C_U(i, j) \\ M(i-1, j+1) + C_R(i, j) \end{cases} \quad (7)$$

$$C_L(i, j) = |I(i, j+1) - I(i, j-1)| + |I(i-1, j) - I(i, j-1)| \quad (8)$$

$$C_U(i, j) = |I(i, j+1) - I(i, j-1)| \quad (9)$$

$$C_R(i, j) = |I(i, j+1) - I(i, j-1)| + |I(i-1, j) - I(i, j+1)| \quad (10)$$

where $P(i, j)$ is an additional pixel based energy measure used on the top of the forward energy cost (e.g., the result of a face detector or supplied by the user in order for specific areas to be protected or removed). In this work the image gradient energy [2] is supplied as this additional pixel based energy measure in order to enhance the results.

During the next step, all the feature vectors produced are passed to an eXtreme Gradient Boosting (XGBoost) classifier [4]. XGBoost is an implementation of gradient boosting machines developed having the performance and computational speed as a goal. It can perform vanilla, stochastic and regularized gradient boosting and it is robust enough to support fine tuning and regularization which, according to its developer, *is what makes it superior and different to other libraries* [4].

IV. EXPERIMENTAL RESULTS

Extensive experiments have been conducted. Four (4) different detectors, Rue [1], Yin [16], Ye [17] and the proposed one, are compared over the Uncompressed Colour Image Database (UCID) [15]. This dataset, which is widely used for image quality assessment, contains 1338 color images with 384×512 or 512×384 pixels/image spatial resolution and its content varies greatly from humans and animals to buildings and landscapes. Images are given in a .tiff format and have never been compressed or preprocessed in any way. All 1,338 images from this dataset are reduced using the seam carving algorithm producing three different sets of images. It must be noted here that the Rubinstein's [3] forward criterion (Eq.7- 10) has been used for resizing the images. In this case $P(i, j)$ is the image gradient energy produced by Sobel's filter.

The first set (called *small set* from this point forward) consists of images reduced from 3% to 21% in steps of 3%. From every different reduction percentage 191 (out of 1338) images were randomly selected and labeled as carved. These 1337 carved images along with the original 1338 form the small set. In the second set (called: *large set*) the images were reduced by seam carving from 10% to 50% in steps of 10%. 267 images from each reduction percentage were randomly chosen in this case. Finally, a *mixed set* created by using all different reduction scales (from small and large set). Images with scaling ratios of 10% and 20% (from the large set) were removed from the mixed set as they were very close to 9% and 21% (from the small set). That means that 1338 original images and 133 randomly selected images for each one of the 10 different ratios are used in the third case.

In order to make fair comparisons all algorithms are developed using the same environment. Python is used as programming language with OpenCV [18] and scikit-learn [19] as image processing and machine learning toolkits, respectively. The experiments were conducted on a 2.3 GHz Intel Core i5 mini-Mac system with 8GB main memory.

A number of linear, non-linear classifiers and ensembles are initially checked to find those who might perform well on the data under investigation. In this phase and for each set 80% of the data were used as training set and the rest 20% for testing purposes. Numerical results (Table II and Figure 3) suggested that XGBoost and SVM classifiers were promising for good output and further investigation. More specific, Table II displays the results for SVM with linear and radial basis function kernels, logistic regression, linear and quadratic discriminative analysis, K-Neighbors and finally decision trees, random forest and XGBoost classifiers. Figure 3 depicts the Area Under the Curve for the two algorithms that gave the best results (to keep the Figure readable).

As mentioned before, SVM classifiers have been used in all previous works. Along with the fact that the eXtreme Gradient Boosting (XGBoost) produces better results in all cases tested made this classifier the one selected for further tuning.

During the next phase, a random search (to reduce the parameter search space) and then an exhaustive grid search

TABLE II
CHECKING DIFFERENT CLASSIFIERS - MIXED IMAGE SET

| Classifier | Train | StDev | Test | AUC |
|------------------------|-------|-------|-------|-------|
| SVM (Linear) | 86.65 | 1.87 | 87.83 | 92.50 |
| SVM (RBF) | 87.58 | 1.95 | 86.70 | 93.52 |
| Logistic Regression | 85.94 | 1.90 | 86.33 | 91.71 |
| Linear Discr. Analysis | 86.65 | 2.37 | 85.02 | 91.93 |
| Quadr. Discr. Analysis | 79.52 | 3.12 | 81.27 | 89.41 |
| KNeighbors Classifier | 83.65 | 2.02 | 81.27 | 88.26 |
| Decision Tree | 84.39 | 3.25 | 81.46 | 81.47 |
| Random Forest | 86.36 | 2.31 | 84.83 | 90.50 |
| XGBoost | 89.17 | 2.09 | 90.07 | 94.93 |

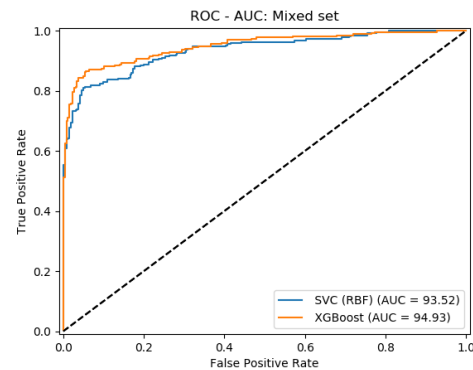


Fig. 3. ROC-AUC: Mixed Set

with 10-fold cross-validation resampling approach was taken for hyper-parameter tuning. The process gave the following set of the hyper-parameters (as they are defined by the scikit-learn python package)

| | |
|--------------------------------|------------------------------|
| $estimators = 500$ | $learning\ rate = 0.025$ |
| $gamma = 0.9$ | $min\ child\ weight = 7$ |
| $objective = binary\ logistic$ | $max\ depth = 17$ |
| $subsample = 0.8$ | $colsample\ by\ level = 0.8$ |
| $lambda = 1$ | $alpha = 0.1$ |

After hyper-parameters tuning an extra step was taken to ensure avoiding overfitting. The train set was divided more in 80% train 20% validation data. The learning process was monitored and an early stop criterion was applied in order to specify the number of the training epochs.

The final model applied to all 3 different test sets and the results are shown in Table III. It is clear that the proposed approach improves the testing accuracy in all cases. Even in the worst case (large set) results are improved from almost 3% to 6%.

In Table III we can also see the results produced by the proposed model when it tries to predict labels for a dataset entirely different from the one used for its training. For this purpose, the RetargetMe dataset [20] is used as an *in the wild* test set. The set includes 80 original images and their seam carved versions under different scaling ratios. The images (content, statistic distributions etc) are entirely different from

those in UCID dataset. In this case, the pipeline proposed herein exhibits 8% to 10% better accuracy compared with the other method even though all the results are significantly lower as expected. For example, the best of the previous proposed methods reports a 58.48% accuracy while the proposed approach 66.08%.

TABLE III
COMPARISON: TEST ACCURACY - DIFFERENT SETS AND METHODS

| Method | Small % | Large % | Mixed % | Retarget % |
|----------|--------------|--------------|--------------|--------------|
| Ryu | 81.46 | 88.48 | 82.83 | 58.48 |
| Yin | 79.81 | 90.47 | 82.96 | 56.14 |
| Ye | 81.31 | 91.4 | 85.21 | 55.56 |
| Proposed | 89.89 | 94.39 | 90.45 | 66.08 |

In a real-life scenarios the scaling ratio of a seam carved image is unknown to the forensics analyst which is why the model for the mixed image set is considered to be the representative one. Table IV and Figure 4 show its classification report and its corresponding confusion matrix.

TABLE IV
CLASSIFICATION REPORT - MIXED IMAGE SET

| | Precision | Recall | F1-score | Support |
|------------------------|-----------|--------|----------|---------|
| | 0.89 | 0.92 | 0.91 | 265 |
| | 0.92 | 0.88 | 0.90 | 269 |
| avg / total | 0.91 | 0.90 | 0.90 | 534 |
| Accuracy 90.45% | | | | |

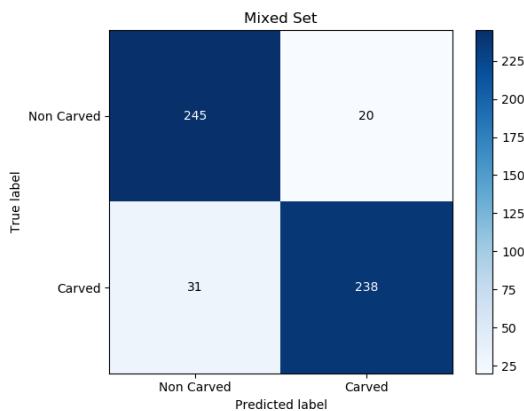


Fig. 4. Confusion Matrix for the Mixed Image Set

V. CONCLUSION

In this paper, a passive-blind approach for detecting seam carved images is investigated. Each feature vector extracted consists of 18 features based on energy (image, seam) and the noise of the image under investigation. All seams related statistics are produced based on the cumulative energy matrix M which is now calculated following the forward energy criterion. An eXtreme Gradient Boosting classifier is trained to automatically determined whether any given image has been

manipulated or not. In this work only the vertical aspect ratio changes are presented. The same procedure can be applied to the horizontal case though. Experimental results based on three different set of images (and on *in the wild*) confirm an improved performance and robustness of the presented approach. An 90.45% test accuracy is reported for the mixed case of seam carved images. As a next step deep learning techniques such as convolutional neural networks should be utilized to further improve the detection results

REFERENCES

- [1] S. J. Ryu, H. Y. Lee, and H. K. Lee, "Detecting trace of seam carving for forensic analysis," *IEICE Transactions on Information and Systems*, vol. E96-D, no. 5, pp. 1304–1311, 2014.
- [2] I. Sobel, *An isotropic 3x3 gradient operator*. Burlington: Academic Press, 1990.
- [3] M. Rubinstein, A. Shamir, and S. Avidan, "Improved seam carving for video retargeting," *ACM SIGGRAPH*, p. 1, 2008.
- [4] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," *CoRR*, vol. abs/1603.02754, 2016.
- [5] T.-T. Ng, S.-F. Chang, C.-Y. Lin, and Q. Sun, *Passive-blind Image Forensics*. Burlington: Academic Press, 2006.
- [6] A. Gallagher, "Detection of linear and cubic interpolation in jpeg compressed images," in *2nd Canadian conference on Computer and Robot Vision*, Washington DC, USA, 2005, pp. 65–72.
- [7] S. Pfennig and M. Kirchner, "Spectral methods to determine the exact scaling factor of resampled digital images," in *5th International Symposium on Communications Control and Signal Processing (ISCCSP)*, Rome, Italy, May 2012, pp. 1–6.
- [8] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM Transactions on Graphics*, vol. 26, no. 3, p. 10, 2007.
- [9] A. Sarkar, L. Nataraj, and B. S. Manjunath, "Detection of seam carving and localization of seam insertions in digital images," in *11th ACM workshop on Multimedia and security - MM&Sec '09*, 2009, p. 107.
- [10] C. Fillion and G. Sharma, "Detecting content adaptive scaling of images for forensic applications," in *Media forensics and security, SPIE proc.*, vol. 75410Z, 2010.
- [11] W. L. Chang, T. K. Shih, and H. H. Hsu, "Detection of seam carving in JPEG images," in *iCAST 2013 and UMEDIA 2013*, Magdeburg, Germany, August 2013, pp. 632–637.
- [12] K. Wattanachote, T. K. Shih, W. L. Chang, and H. H. Chang, "Tamper Detection of JPEG Image Due to Seam Modifications," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 12, pp. 2477–2491, 2015.
- [13] Q. Liu, "Exposing seam carving forgery under recompression attacks by hybrid large feature mining," in *International Conference on Pattern Recognition*, 2017, pp. 1041–1046.
- [14] J. D. Wei, Y. J. Lin, and Y. J. Wu, "A patch analysis method to detect seam carved images," *Pattern Recognition Letters*, vol. 36, no. 1, pp. 100–106, 2014.
- [15] G. Schaefer and M. Stich, "UCID: an uncompressed color image database," in *Proc.SPIE*, vol. 5307, 2003, pp. 472–480.
- [16] T. Yin, G. Yang, L. Li, D. Zhang, and X. Sun, "Detecting seam carving based image resizing using local binary patterns," *Computers and Security*, vol. 55, pp. 130–141, 2015.
- [17] J. Ye and Y.-Q. Shi, "A Local Derivative Pattern Based Image Forensic Framework for Seam Carving Detection," in *16th International Workshop on Digital Forensics and Watermarking, IWDW 2017*, Magdeburg, Germany, August 2017, pp. 172–184.
- [18] OpenCV (open source computer vision library). Accessed: 2018-04-20. [Online]. Available: <https://opencv.org>
- [19] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [20] M. Rubinstein, D. Gutierrez, O. Sorkine, and A. Shamir, "A comparative study of image retargeting," *ACM Transactions on Graphics*, vol. 29, no. 6, p. 1, 2010.

SIMULATION FOR DATA SECURITY IMPROVEMENT IN EXPLOITED METAROUTER

¹Kristono, ²Imam Riadi

¹Department of Informatics, Universitas Islam Indonesia

²Department of Information System, Universitas Ahmad Dahlan

Email: ¹ kristonosukses58@gmail.com, ²imam.riadi@is.uad.ac.id

Abstract— Implementation of data security used in this research using metarouter as its method. Metarouter is a virtual network device that connects computers as if in a network. Metarouter is made to make it easier to monitor network activity simultaneously. This study aims to develop data security management on metarouter. Testing conducted by Denial of Service attacks based on DOS flooding metarouter aimed at port 80 and port 22. To recognize Denial of Service attacks it is necessary to monitor the network by performing log analysis stored on mikrotik. The benefits of log analysis are expected to facilitate in data monitoring and network management.

Keywords: *Metarouter, DoS (Denial of Service), Data Security, Mikrotik*

I. INTRODUCTION

The computer network is a collection of several computers connected together via a wired or wireless and can communicate with one another by using rules (protocol) specific. Managing a network consisting of multiple computers is work still to be done easily. However, if the network is growing, then to manage the network will be very difficult for any network manager [1].

To manage such a large scale network with the network (network) it should be separated into several smaller networks. Set some small network containing dozens of hosts, it would be easier than arranging a network comprised of hundreds or even thousands of hosts. Mechanical separating these networks can be implemented on a network (LAN), medium-scale network (MAN) or large networks (WAN / Internet) [2].

Once the network is separated into several smaller networks, the next job is to reconnect the small networks. In the network topology in a lab has room for Practical, Server, technicians, and lecturers. Each room has a need and Access Control Lists (ACLs) are different. ACLs on Computer Laboratory Computer Network centered on the router server and centralized ACLs that many can lead to congested traffic. Separation router ACLs impact on the use of more and cause excessive cost to purchase a router, power consumption and the use of storage space. Those problems can be solved by virtualization. Mikrotik Router can implement virtualization with Metarouter that impact on the cost savings of purchasing hardware router, electricity usage, and storage. Virtualization routers use

Metarouter can save the cost of making a computer network, electrical energy consumption and the use of space than non-virtualized routers [3].

The router is an important device in a network, a lot of evidence that can be drawn from the activities of the network, in addition to the router is also intelligently able to know where the flow of information purposes (quotas) to be passed. The evidence can be drawn from the routers include firewall configuration, mac address, IP address client list, activity logging and other admin [4].

RouterOS is the operating system and software that can be used to make the ordinary computer into a reliable network router, includes a variety of features that are made for IP networks and wireless networks. These features include Firewall & Nat, Routing, Hotspot, Point to Point Tunneling Protocol, DNS server, DHCP server, Hotspot, and many other features [5].

The proxy can be used in two types, namely in the form of hardware and software. In the form of hardware, Mikrotik usually already installed on a particular board, whereas in the form of software, Mikrotik is a Linux distribution that is dedicated to the function of the router. MikroTik RouterOS™ is the Linux operating system base is intended as a network router. Designed to suit all users. The administration can be done through the Windows Application (WinBox). Besides the installation can be done on a Standard PC (Personal Computer). PC which will be used as a proxy router also does not require resource large enough to use standard, for example, only as a gateway. For the purposes of a large load (a complex network [6],

Metarouter Mikrotik is a feature that allows running the new operating system in a virtual good for application virtualization and virtualization router network topology. Almost the same as VMware or VirtualPC application. With Metarouter a proxy Routerboard will be able to run some sort of virtualization apart RouterOS Router OS with Metarouter can also run an OS other OpenWRT Linux operating system instance. For that to Metarouter allowed in a single router can be used for various things such build RouterOS Virtual.

Virtual Server build, also can build a network topology. Also can be used to simplify the configuration which when put together will be very difficult or even confusing, for example,

for load balancing two ISPs at the same bandwidth at the same manager also coupled VPN firewall, it would be wise to be separated and run on Metarouter in an h / w Mikrotik. From usability very much Metarouter allows us more efficient than the use of power to the router as if going to have a lot of routers are used as well as more practical [7]. Network topology on a meta-router described in Figure 1.

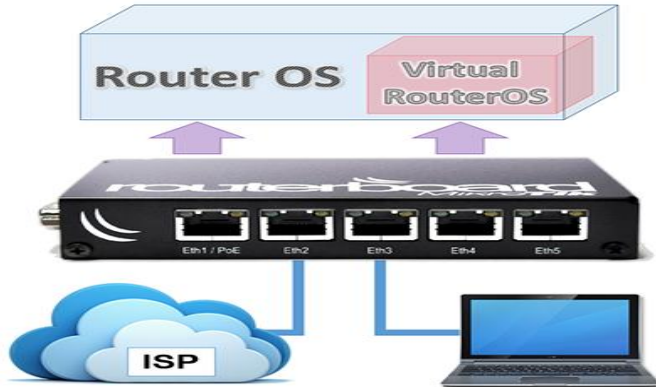


Figure 1. Topology Metarouter

Using the topology Metarouter as in Figure 1, Metarouter we build more or less like the image above so that the client as though it has its own router. And we as Admins, we can still be a physical router management. Before building a virtual machine, we need to determine in advance a big RAM and hard drive which will be allocated to the virtual router. Mikrotik Operating System, with at least 24 MB of RAM is recommended. For the size of the hard drive can be customized to your needs. If the last is already defined then it is time we run virtualization in MikroTik router Metarouter feature.

MikroTik operating system is designed as a network router. And which can be used to make the computer into a reliable network router. The function of MikroTik include Firewall & Nat, Bandwidth Limiter, Routing, Hotspot, Point to Point Tunneling Protocol, DNS Server, DHCP Server, Hotspot and more a function of MikroTik [8].

Connected to a network, the computer is vulnerable to infiltration from outside. If someone can infiltrate a computer then the person can take the data that is stored on your computer and use it for personal gain. Data security become important in data communication is done. When the data user ID and password of the service that we use fall into the wrong hands, it could be that people will use for things that are not responsible. Data security is an activity to keep the resource information remains secure. It is required in a network activity monitoring is required in order to access path made suspicious data access can be resolved before things refrigerated not happen [9].

Based on the background that has been described in the realm of this study is to exploitation and digital monitoring contained in the proxy Router OS by utilizing Metarouter as the media in the implementation of which is used for data security with simulation methods. Where computers connected in a network as if have router itself in the management of their networks, with Metarouter which has made it easier for traffic monitoring user activity without disturbing other users, although in one

Routerboard. Metarouter also allows monitoring of multiple activities simultaneously without the user using only one Routerboard [10].

II. LITERATURE REVIEW

a. Network Management

Network management ability is to control and monitor a computer network from a location. The International Organization for Standardization (ISO) defines a conceptual model to explain the function of network management. Fault Management, provides a facility which allows the network administrator to find out the fault on a managed device, network, and network operations, in order to immediate determination what the cause is and can immediately take action.

Fault management mechanism for Reporting of errors, logging, diagnosing, Correc errors ,Configuration Management, monitors the network configuration information so that the impact of any hardware or specific software can be managed properly. This can be done with the ability to initialization, reconfiguration, deployment, and off-managed devices [11].

Performance Management, is to measure various aspects of network performance including the collection and analysis of statistical data so that the system can be managed and maintained at a certain level that is acceptable.

Performance management has the ability to Obtain utilization and error rates of network devices, Maintaining a certain level of performance to ensure data card has enough space. Security Management, manage access to network resources so that the information can not be obtained without permission. This is done by limiting access to network resources, give notice of the existence of business violations and security breaches. The groove security management describes in Figure 2.

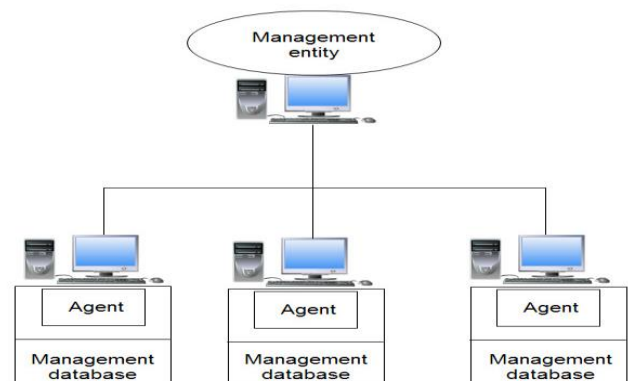


Figure 2. Network Management Architecture

In Figure 2 describes Network Management Station (NMS), running a network management application that is able to gather information about the managed device from the management agent located in the device. Network management applications have to process large amounts of data, react to certain events,

and prepare the relevant information to be displayed. NMS usually has a control console with a GUI interface that allows users to view a graphical representation of the network, control devices in the network are managed and programmed the network management application. Some network management applications can be programmed to react to the information obtained from the management agent and / or set the threshold value in order to Conducting tests and automatic correction, logging Provides status and warning information to the user.

Managed devices, such as all kinds of devices in the network, such as a computer, printer, or router. In the device, there is a management agent. Management agents, provide information about managed devices to the NMS, and may also receive control information/control. Network management protocol, used by the NMS and the management agent to exchange information. Management information, the information exchanged between the NMS and the management agent that allows the monitoring and control of the device. be any type of device in the network, such as a computer, printer, or router. In the device, there is a management agent. Management agents, provide information about managed devices to the NMS, and may also receive control information/control.

Network management protocol, used by the NMS and the management agent to exchange information. Management information, the information exchanged between the NMS and the management agent that allows the monitoring and control of the device. be any type of device in the network, such as a computer, printer, or router. In the device, there is a management agent. Management agents, provide information about managed devices to the NMS, and may also receive control information/control. Network management protocol, used by the NMS and the management agent to exchange information. Management information, the information exchanged between the NMS and the management agent that allows the monitoring and control of the device [12].

Network management software (network management applications and agents) are usually based on a specific network management protocol and network management capabilities provided by the software is usually based on the functions that are supported by a network management protocol. Network management software selection is determined by the network environment (the range and nature of the network), network management requirements, fees, operating system. Network management protocol most commonly used is the Simple Network Management Protocol (SNMP), Common Management Information Protocol (CMIP). SNMP is a protocol that is most widely used in the local network environment (LAN). Meanwhile, CMIP is used in the telecommunications environment, where larger and more complex networks) [13].

b. Router

Basis Technology Router is a device that sends data packets through a network or the Internet to the destination, through a process called routing. The router serves as a liaison between two or more networks to carry data from one network to another. Either the same or different networks in terms of

technology as connecting a network that uses a bus topology, Star, and Ring. Such small networks into a larger network, called the internetwork, or to divide a large network into several subnetworks to improve performance and simplify management.

Routers main function is to route packets. A router has routing capabilities, Router intelligently means to know where the travel route information (package) will be missed, whether intended for other hosts that the network or is on a different network. If packets addressed to a host on another network router will forward it to the network. Conversely, if the packets addressed to host a network router will block the packets out. Figure 1 shows an example of a network with the network [14].

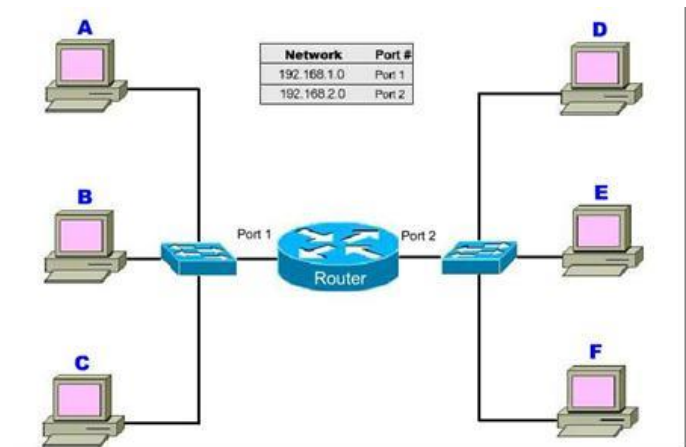


Figure 3. Example of a Network to Network

In Figure 3 there is two network connected to a router. Network left connected to port 1 routers have a network address 192.168.1.0 and the network to the right is connected to port 2 of the router with network address 192.155.2.0.

Computer A sends data to computer B, then the router will not forward the data to another network. 2. Similarly, when the computer sends data to DF, the router will not forward data packets to another network. 3. Only when the F transmit data to computer B, then the router will continue data packet to computer B Mikrotik is a computer operating system and computer software that is used to make ordinary computer into a router, proxy divided into two proxy system operation can be worn Mikrotik os and Mikrotik board, to Mikrotik board does not require a computer to run it enough to use board that already include the proxy os.

Mikrotik OS includes features created specifically for IP networks and wireless networks. Mikrotik operating system is the base Linux operating system that is used as a network router. created to provide convenience and freedom for its users. The administration settings can be done using the Windows Application (WinBox). The computer that will be used as a proxy router also does not require a high specification, for example, only as a gateway. Unless the proxy used for large loads (complex networks, routing complex) should use adequate specifications.

c. Mikrotik Router OS

Mikrotik is an operating system includes software that is installed on a computer so that the computer can act as the heart of the network, controlling or regulating data traffic between networks, this type of computers known as routers. So basically Mikrotik is one of the special operating systems for the router. Mikrotik Router is known as one of the OS that is reliable and has a lot of features to support a smooth network. Mikrotik Router can be used on small scale computer networks or large, it is, of course, adapted to the resource rather than the computer itself. If the proxy is used to set up a small network can then use the computer device that is mediocre [15],

Mikrotik RouterOS is the operating system Mikrotik RouterBOARD hardware. It can also be installed on the PC and will turn it into a router with all the necessary features - routing, firewall, bandwidth management, wireless access points, backhaul link, hotspot gateway, VPN server and much more. RouterOS supports various configuration methods - local access to the keyboard and monitor, serial console with a terminal application, access Telnet and SSH secure through the network, configuration tools custom GUI called Winbox, the interface is Web-based configuration is simple and the programming interface API for building control applications you itself If there is no local access, and there is a problem with the IP level communications, RouterOS also supports MAC level based connection with Mac-Telnet tool and Winbox specially made.

- Winbox GUI over IP and MAC
- CLI with Telnet, SSH, local console and console serial
- API to program your own tool
- The web interface

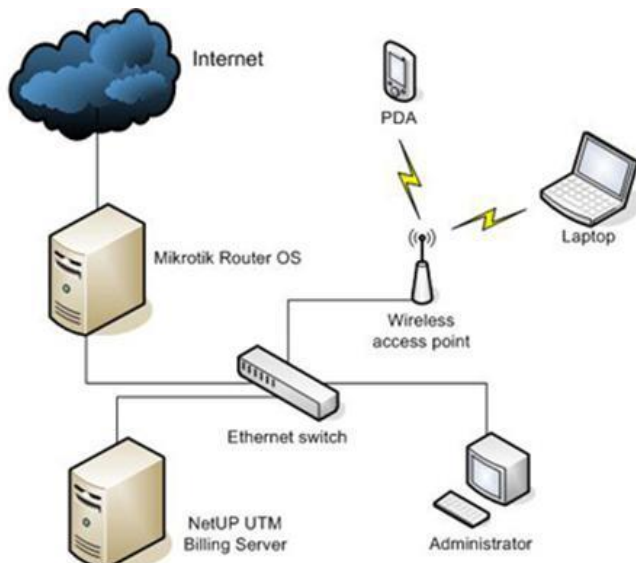


Figure 4. Use of Mikrotik Router OS

On Figure 4 describes scheme IP address allocation via DHCP, authorization page, and Internet access switching on/off is performed by a router operated by Mikrotik RouterOS (hereafter Mikrotik). Tariffs and user database is stored in the

billing system on a separate server. Authorization and accounting are processed via RADIUS protocol.

d. Metarouter

Metarouter Mikrotik is a feature that allows running the new operating system virtually. Almost the same as VMware or VirtualPC application on Windows. Metarouter we can use to run on OS Mikrotik operating system is running. By using Metarouter, a client as if it has its own router. And we as administrators, we still can management physical router. Before building a virtual machine, we need to specify in advance a large RAM and hard drive that will be allocated to the virtual router. With Mikrotik Operating System, the suggested minimum RAM is 24MB. For the size of the hard drive can be adjusted as needed. If the parameters had been determined then it is time we run the virtualization in the Mikrotik router with features Metarouter [16][17].

Metarouter has some limitations as follows :

1. Just can run up to eight (8) virtual machine for each RouterBoard.
2. Unable to use CF or MicroSD
3. Sometimes OpenWRT not to-shutdown perfectly on time
4. RouterOS experience Reboot.
5. Virtual RouterOS cannot use the wireless interface that is owned by the RouterBoard.

Benefits of Using Meta Router Router Virtualization

1. Virtualization can be applied to build several virtual RouterOS.
2. Virtualization buffer applied to build a virtual server, such as Web
3. Server, FTP Server, DNS Server, Database Server, VoIP Server, Proxy Server and others.
4. Virtualization can be applied to build a network topology.
5. Virtualization with Metarouter will not burden the PC or Laptop, it is because Metarouter runs on RouterBoard.
6. Virtualization will be more concise and compact as it can be packed in a casing RouterBoard. This brings advantages for network
7. Your virtual can be easily carried loaned you. Very useful for trainers or instructors who often conduct training and presentations in different places.
8. More power-efficient, because it only requires a portion of a RouterBoard to get eight (8) units RouterOS
9. The operating system used by the virtual RouterOS operating system is legal.

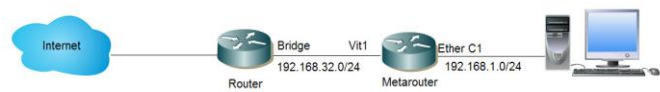


Figure 5. Utilization of Metarouter

On the Figure 5 explain about the benefits of virtualization can be applied i.e. meta-router to build several virtual RouterOS,

virtualization encompasses applied to build virtual servers, such as Web, virtualization can be applied to build network topology and virtualization with Metarouter won't overload your PC or Laptop, this is because Metarouter run on Routerboard.

e. Data security

The data transmitted over the Internet network is the most important data. It invites others to steal and exploit these data for their own interests. Certainly would be detrimental to the owner of the data. Theft and use of data by unauthorized persons is a crime. Internet plays an important role in human life today. Many of the activities carried out by utilizing the internet. These data sent from the user's computer to the server computer service provider used. Before arriving at the server computer service provider, the data transmitted will pass through the computers on the Internet. At the time of passing through the Internet network, data are transmitted vulnerable to eavesdropping. In addition to tapping. Computers that are used can be infected by a virus that works as spyware.

Where spyware can record the activities carried false. Connected to a network the Internet, the computer is vulnerable to infiltration from outside. If someone can infiltrate a computer then the person can take the data that is stored on your computer and use it for personal gain. Data security become important in data communication is done. When the data user ID and password of the service that we use fall into the wrong hands, it could be that people will use for things that are not responsible. then a computer is vulnerable to infiltration from outside.

If someone can infiltrate a computer then the person can take the data that is stored on your computer and use it for personal gain. Data security become important in data communication is done. When the data user ID and password of the service that we use fall into the wrong hands, it could be that people will use for things that are not responsible. then a computer is vulnerable to infiltration from outside. If someone can infiltrate a computer then the person can take the data that is stored on your computer and use it for personal gain. Data security become important in data communication is done. When the data user ID and password of the service that we use fall into the wrong hands, it could be that people will use for things that are not responsible [18][19].

Based on the research results, there is no computer network completely safe from hackers, crackers, spam, e-mail bombs, computer viruses etc. What you can do is keep from the network easily penetrated, while continuing to try to improve the system of data and network security. In the current global era, the Internet-based information system security is a must for the more, because of the public Internet network and global nature inherently unsafe.

At the time of data sent from one computer to another on the Internet, the data will pass through a number of the different computer. Means will provide an opportunity to the user to take over one or multiple computers. Unless a computer is locked nature of a room that has limited access to the outside of the room, then the computer will be safe. Burglary security system on the Internet occurs almost every day throughout the world.

Cybercrime or better known as Cyber Crime is a form of virtual crime by the media using a computer connected to the Internet and exploit other computers connected also to the Internet.

The holes in the operating system causing weakness and the opening hole that could be used by hackers, crackers and script kiddies to infiltrate into the computer. Crimes that occur can be theft of data, access to the internal network, changes to important data and information theft resulted in the sale of information. Cybercrime or better known as Cyber Crime is a form of virtual crime by the media using a computer connected to the Internet and exploit other computers connected also to the Internet. The holes in the operating system causing weakness and the opening hole that could be used by hackers, crackers and script kiddies to infiltrate into the computer.

Crimes that occur can be theft of data, access to the internal network, changes to important data and information theft resulted in the sale of information. Cybercrime or better known as Cyber Crime is a form of virtual crime by the media using a computer connected to the Internet and exploit other computers connected also to the Internet. The holes in the operating system causing weakness and the opening hole that could be used by hackers, crackers and script kiddies to infiltrate into the computer. Crimes that occur can be theft of data, access to the internal network, changes to important data and information theft resulted in the sale of information.

The holes in the operating system causing weakness and the opening hole that could be used by hackers, crackers and script kiddies to infiltrate into the computer. Crimes that occur can be theft of data, access to the internal network, changes to important data and information theft resulted in the sale of information. The holes in the operating system causing weakness and the opening hole that could be used by hackers, crackers and script kiddies to infiltrate into the computer. Crimes that occur can be theft of data, access to the internal network, changes to important data and information theft resulted in the sale of information.

f. WinBox 3.8

Winbox is a software or utility that is used to remotely a proxy server mode in GUI (Graphical User Interface) through the Windows operating system. People are more configure proxy or proxy Routerboard os using Winbox compared with the configured directly through mode CLI (Command Line Interface). It was caused, not least because the process is more simple and easy and by using this. Winbox software configuration server can be completed quickly compared with the CLI mode have should memorize and type the proxy console [20].

Winbox main function is to exist in the proxy settings, it means that the main task is to set its window or set a proxy with GUI, or the desktop. Winbox more detailed function is

1. Setting Mikrotik router
2. Limit Setting network bandwidth
3. for setting block a site
4. Setting Hotspot Login

5. Setting network security

Configure a Mikrotik Winbox via it's more widely used because in addition to its easy-we also don't have to recite the commandments console.

III. IMPLEMENTATION

This section includes the results obtained during the research that has been done by the formulation and purpose of the study, namely: 1) Setting the meta-router and data security 2) to simulate meta exploration router follow figure 6

A. IP network settings

Before setting up a virtual router on this section, we first perform IP settings that we will use to the client which will be connected to our proxy. We determine the IP that we will make as a public IP address we give to Client. Here's how the IP network settings on the Routerboard Mikrotik. These stages are listed in Figure 6.

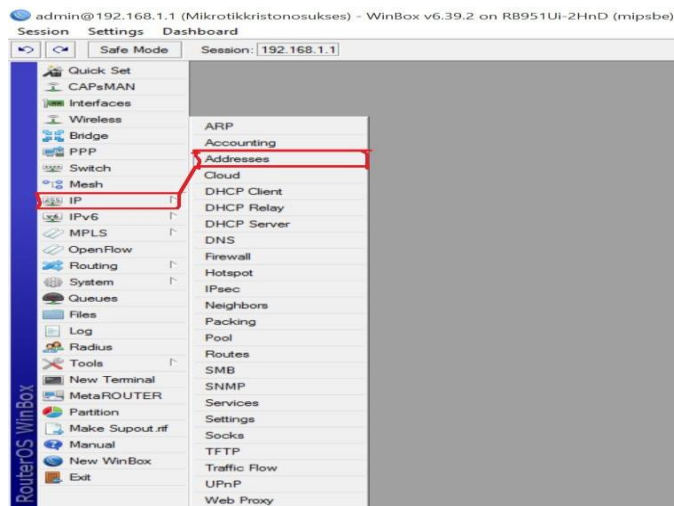


Figure 6. Setting IP Network in Virtual Box

Figure 6 describes the first step in the route the IP network settings in virtual box. Input the IP address and the subnet enter the classmate with locally input router default gateway i.e. IP address belongs to the router and DNS IP address that is public, and then click ok.

Later in the virtual box, we can set the IP, the determination and the determination of classification class IP. These stages are listed in Figure 7.

In this section explains how to make settings meta-router steps we can do First, go to Metarouter. Click the + button to add a virtual router. Here there are three parameters that need to be determined, "Name" is filled with the name of the Virtual Router suit your needs

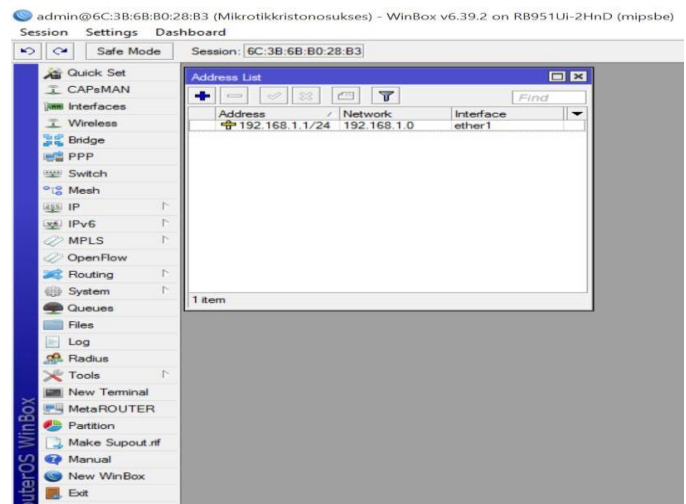


Figure 7. Generation of IP Address

Parameter RAM and hard drive then also filled in regarding the needs. Other parameters can be left worth defaults. These stages are listed in Figure 8.

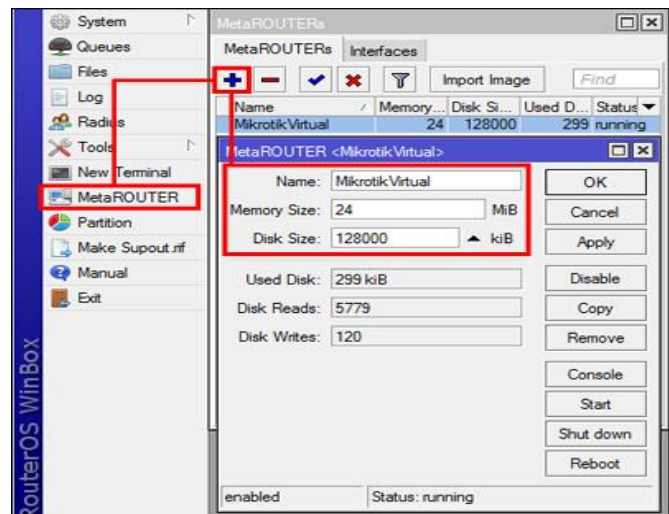


Figure 8. Creating Metarouter

Once we "Apply" Automatic Virtual Router in Metarouter will run. This step in accordance with that listed in Figure 8. Operating System will automatically use Mikrotik RouterOS and run the same version with Router Mikrotik RouterOS version. Virtual Router does not yet have an ethernet interface, and we can not communicate in a network with other devices, but can only be accessed by the console. To access the router via a virtual console, right-click the virtual router on Metarouter menu and then select the "Console", or also via a terminal with the command: / meta-router console [virtual-router-name].

Figure 9 describes how to access the virtual router via console, right-click the virtual router on the menu then choose Metarouter "Console ", or via the terminal with the command:/Metarouter console [name-virtual-router]

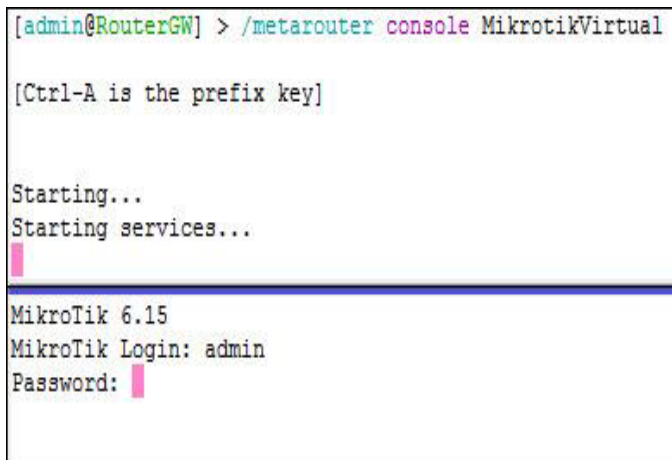


Figure 9. Meta Console Router

The next step we need to do is create a virtual ethernet. The virtual ethernet which will be used by the virtual router to be able to communicate with the master router or even other devices in the network. How to make virtual ethernet, enter the IP menu -> Click the + (add). Then select Virtual Ethernet.

In Figure 10 describes how to make setting virtual ethernet which later will be used by the virtual router to be able to communicate with the router is master or even other devices in

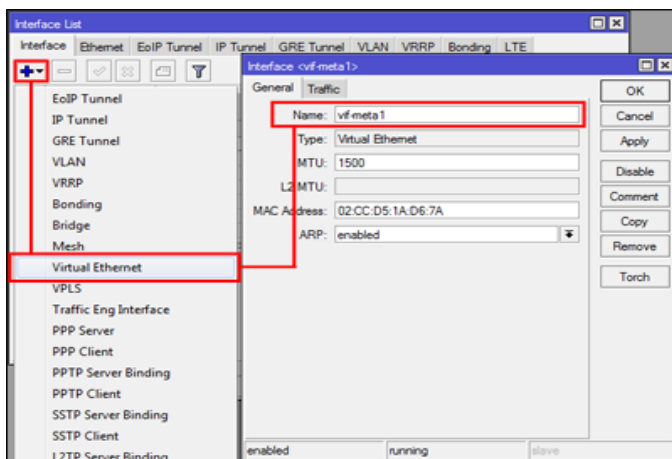


Figure 10. Setting the Virtual Ethernet

the network. How to make a virtual ethernet, IP-enter the menu-button click > + (add). Then select Virtual Ethernet.

One virtual ether will be used by the virtual router to be able to communicate with the Router Master, and another one virtual ethernet router for virtual communication with the hosts in the network, such as a laptop client. If it is, we define virtual ethernet to the router so that it can be used by the virtual router. Sign into Metarouter menu, then click on the tab "Interface". Click Button + (add). These stages are listed in Figure 11. Figure 11 describes how to make setting meta-router. In the "Virtual Machine" in the virtual select which virtual ethernet router will be used. Then in the "Type", select the "static".

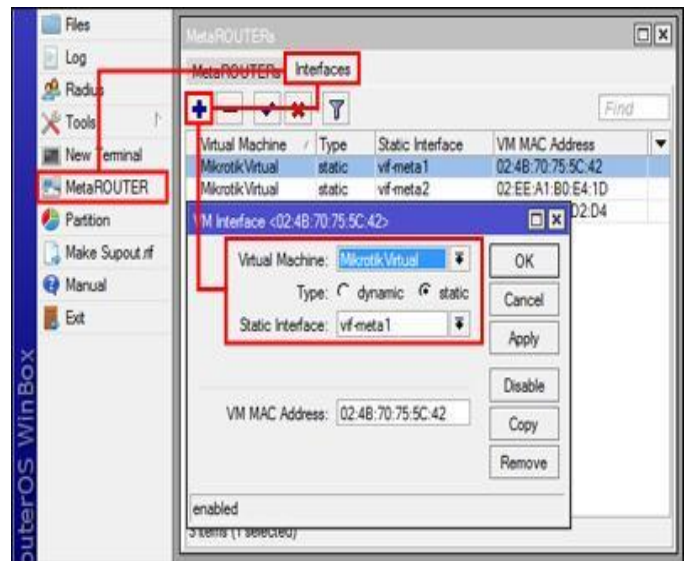


Figure 11. Interface Setting Metarouter

Finally, the option "Static Interface", select the virtual ethernet previously made. In the example above, we will be setting up a virtual router with two Ethernet interfaces. One for communication virtual ethernet router to the Internet, one for communication to the client. To be sure, try the virtual remote router console and display the ethernet interface.

The ether1 virtual interface in the router can we communicate with the interface on the master router, how pretty setting the IP address of the segment between the virtual interfaces on the router master with the ether1 interface in the virtual router. While virtual router ether2, still can not communicate with other devices or client, in order to communicate we need to bridge the physical interface connected to a network client.

B. Data Retrieval Metarouter

This section will explain the steps Pengambilan metadata on the router to perform the interconnection between the ether that we have become a virtual router setting at this stage of the above settings.

The next step is to connect the data collection in meta-router. First, do the IP settings on the client computer that will do PenTest. The first step makes the client meta-router. After we make our next client meta-router to connect to client meta-router After the new interconnected Client we make settings capture data packets between clients in the meta-router.

C. Scenario Testing DoS attacks (Denial of Service)

DoS (Denial of Service) is a type of attack on a computer or server in the internet network by spending resources (resource) owned by that computer until the computer is not able to function properly thus indirectly prevent other users to gain access to services from the attacked computer. This scenario listed in Figure 12.

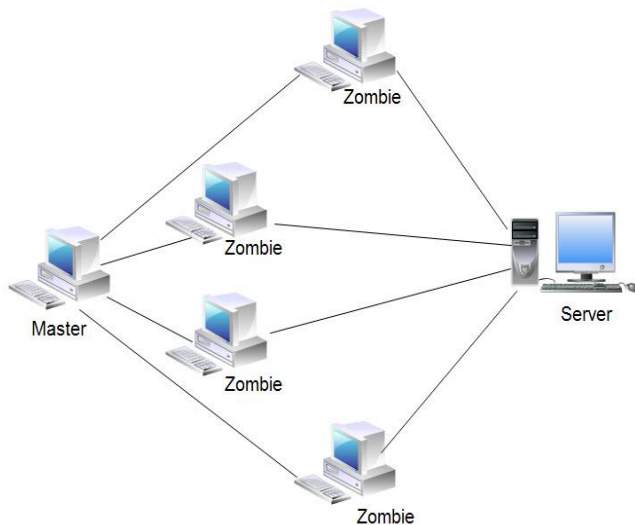


Figure 12. The Concept of Dos Attack

In Figure 12 describes a Denial of Service attacks, the attacker will try to prevent a user access to a system or network using a number of ways, which is as follows:

- Flooding the network traffic with a lot of data so that network traffic coming from unregistered users will be unable to enter into the network system. This technique is referred to as traffic flooding.
- Flooding the network with a lot of requests to a network service in the provided by a host so that requests are coming from a registered user cannot be served by such services. This technique is referred to as a request flooding.
- Interfere with communication between a host and a client that is registered by using a lot of ways, including by changing the system configuration information or even the physical destruction of the components and servers.

Forms of Denial of Service attacks are attacks initial SYN Flooding Attack, which first appeared in 1996 and exploit the weaknesses contained in the protocol Transmission Control Protocol (TCP). Other attacks eventually developed to exploit vulnerabilities present in the operating system, network services or applications to make the systems, network services, or applications cannot serve the user or even crash. Some of the tools used to perform DoS attacks was developed after it (even some freely available tools). Including Bonk, LAND, Smurf, Snork, WinNuke, and Teardrop.

Figure 13 describes the DOS attack by using ICMP ping use. This attack will make the down server. DOS attacks done by flooding a site or the servers with a lot of traffic or packet data until the server is unable to process all requests in real time or concurrently and finally down or paralyzed.

Nevertheless, attacks on a TCP DoS attacks are often performed. This is because the other type of attack (as well as filled the hard disk in the system, locking one valid user account, or modify the routing table in a router) requires penetration of the network first, that the possibility of penetration is small, especially if the network system has been reinforced.

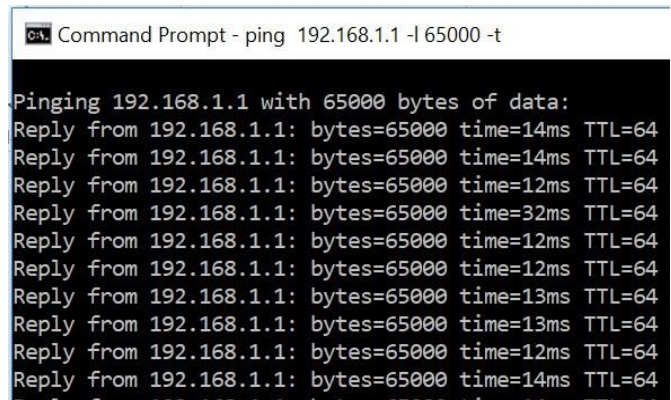


Figure 13. Simulated DOS Attack with ICMP Ping

Denial of Service attack is a hacking technique to make down or paralyzed site or server by flooding the site or server with a lot of traffic or packet data so that the server can not process all requests in real time or concurrently and finally down or paralyzed.

The state of the traffic before the attack showed the data memory and CPU significant move yet transaction the DoS will affected the performance of the proxy where the presence of Dos attacks CPU will experience a significant increase in this case access will be causing the performance of the data packets to be down.

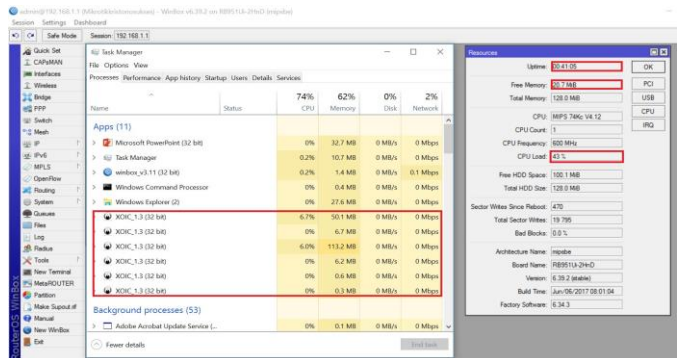


Figure 14. Result DOS Attack

Figure 14 describes explain about the results of a DOS attack against the Server before the server so that the safeguards be done down, CPU load increase.

D. Data security

DoS attacks (Denial of Service) can cause overloading router. Which means that the CPU usage reaches 100% and routers can be unaffordable to timeout. All operations on the package can use significant CPU power as a firewall (filters, NAT, mangle), logging, queues can cause overloading if too many packets per second arrive at a router.

Generally, there is no perfect solution to protect against DoS attacks. Each service can be overloaded because of too many requests. But there are some methods to minimize the impact of attacks.

The first step is to do is Doing Diagnosis attacks on network traffic.

- Connecting diagnosis Firewall
- Doing Diagnosis connection on The network interface
- Perform diagnosis on the CPU performance

After we make the diagnosis of network traffic, which we will evaluate is the traffic network, whether the network is secure or there are symptoms of an attack from the outside. In this case, we can find out suspicious traffic activity, the aim is to prevent damage to data and anticipation of attacks that are not cooled so that we can protect and prevent against attacks that damage the data with no restrictions on access.

IV. CONCLUSION

The conclusion that has been obtained during the research process in the Simulation For Improved Data Security At Metarouter Already Exploited Children concluded that: Metarouter besides more efficient in terms of financial, also easy to perform network management and security protection. Diamana only with the security settings will affect all clients are created in meta-router, while also facilitating the monitoring network that can attack resolved

quickly and easily. Metarouter Allows for data security and data log retrieval is done in the development of a panel Routerboard although there are a few clients in the management and customized traffic on client management division

V. FUTURE WORK

Metarouter only contained in one Routerboard that cannot be updated automatically. This is overcome when there is damage to RouterBoard physically. To understand Metarouter so as not to trouble, to distinguish which router which virtual and physical, we could set the System Identity of each router. Since virtual router interfaces do not yet have, we need to do is create a virtual ethernet.

REFERENCES

- [1] Albert, S., & Juni, E. (2015). Security System Analysis Of Data Network Based On VPN. *Stmik Ikmi*, 10(18), 220.
- [2] Fahri, M., Fiade, A., & Suseno, H. B. (2017). Network Simulation Of Virtual Local Area Network (VLAN) Using Pox Controller. *Journal Of Informatics Engineering*, 10(1), 85–90.
- [3] Galang, C. M., Eko, S., & Imam, A. (2017). Virtualization Techniques Using Router Mikrotik Metarouter (Case Study: Computer Network Laboratory Politeknik Negeri Lampung), 2641–2644.
- [4] Mazdadi, M. I., Riadi, I., & Luthfi, A. (2017). Live Forensics on RouterOS using API Services to Investigate Network Attacks. *International Journal of Computer Science and Information Security (IJCSIS)*, 15(2), 406–410.
- [5] Riadi, I. (2011). Optimization Of Network Security Using Application Filtering Based Mikrotik Introduction Foundation Of Theory. *JUSI, Universitas Ahmad Dahlan Yogyakarta*, 1(1), 71–80.
- [6] Ghazali, T., & Indriati, K. (2016). Network Simulation Of Multi-Protocol Label Switching And Traffic Engineering. *Jurnal Elektro*, 9(1), 23–34
- [7] Soon, J. N. P., Abdulla, S. H. R., Yin, C. P., Wan, W. S., Yuen, P. K., & Heng, L. E. (2013). Implementing of Virtual Router Redundancy Protocol in a Private University. *Journal of Industrial and Intelligent Information*, 1(4), 255–259.
- [8] Bhayangkara, F. J., & Riadi, I. (2014). Implementation Proxy servers and Load Balancing using the method Per Connection (PCC) Classifier Based Mikrotik (Studi case : Shmily.net). *Sarjana Teknik Informatika*, 2, 1206–1217.
- [9] Yudhistira, D. S., Riadi, I., & Prayudi, Y. (2018). Live Forensics Analysis Method For Random Access Memory On Laptop Devices, *IJCSIS) International Journal of Computer Science and Information Security* 16(4).
- [10] Fietyata, Y., & Prayudi, Y. (2013). Technique Exploration Of Digital Evidence On File Sharing SMB Protocol To Support Digital Forensic On Computer Networks. *National Informatics Conference*.
- [11] Gao, X., Zhang, X., Lu, Z., & Ma, S. (2013). A General Model for the Virtual Router. In *A General Model for the Virtual Router* (pp. 334–339). ICCT.
- [12] Mualfah, D., & Riadi, I. (2017). Network Forensics For Detecting Flooding Attack On Web Server. *IJCSIS) International Journal of Computer Science and Information Security*, 15(2), 326–331.
- [13] Komang, I. G., Mardiyana, O., Komang, I. G., & Mardiyana, O. (2015). Network Security With Firewall Filter Based Mikrotik On Computer Lab STIKOM Bali. *Stmik Stikom*, 1(86), 9–10.
- [14] Arif Hidayat. S.T., M. K. (2016). Building A Expert System Application To Help Problem Solving Network On Mikrotik Router. *Jurnal Mikrotik*, 6.
- [15] Riadi, I., Eko, J., Ashari, A., & -, S. (2013). Internet Forensics Framework Based-on Clustering. *International Journal of Advanced Computer Science and Applications*, 4(12), 115–123.
- [16] Luthfi, A., Prastya, S. E., & Luthfi, A. (2017). Forensic Analysis of Unmanned Aerial Vehicle to Obtain GPS Log Data as Digital Evidence Forensic Analysis of Unmanned Aerial Vehicle to Obtain GPS Log Data as Digital Evidence. *IJCSIS) International Journal of Computer Science and Information Security*, 15(April), 280–285.

- [17] Zulkifli, M. A., & Dahlan, U. A. (2018). Live Forensics Method for Analysis Denial of Service (DOS) Attack on Routerboard. *International Journal of Computer Applications*, 180(35), 23–30.
- [18] Amirullah, A., Riadi, I., & Luthfi, A. (2016). Forensics Analysis from Cloud Storage Client Application on Proprietary Operating System. *Ijca*, 143.(June), 1–7
- [19] Isnanto, R., & Diponegoro, U. (2017). The Development Of Computer Network Security System Based On Network Forensics Analysis. *Journal Of Computer Science Electrical Engineering And Informatics*, 3, 12–19.
- [20] Isnanto, R., & Diponegoro, U. (2012). Diponegoro University Network Simulation With Multiprotocol Label Switching, (September 2015).

Optimization Algorithm to Control Interference-based Topology Control for Delay-Constrained Mobile Ad-hoc Networks

Selvakumar¹, S. Selvakumar²

¹ Research Scholar, Department of Computer Science and Engineering, SRM Institute of Science and Technology, E-mail id: selvakumar1985s@rediffmail.com

² Assistant Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, E-mail id: selvakumar.s@ktr.srmuniv.ac.in

Abstract—To increase the network capacity, there is need to minimize the interference among nodes and optimum control of topology in the foundation of network. Recently, technological development helps to build of mobile ad-hoc networks (MANETs) in order to improve the quality of service (QoS) in terms of delay. In contradictory to the objective of minimizing interference, it is important to concern topology control in delay constrained environment. The present research work attempts to control the delay-constrained topology with jointly considering delay and interference concept. Additionally, the study proposed an interference oriented topology control algorithm for delay-constrained MANETs by taking account of both the interference constraint and the delay constraint under the specific condition of transmission delay, contention delay and the queuing delay. Further, the study investigated the impact of node mobility on the interference oriented topology control algorithm. Finally, the results of the present study shows that the proposed algorithm controls the topology to convince the interference constraint, and increases the transmit range to congregate the delay requirement. Also, the study conclude that the algorithm could effectively reduce the delay protocol and improve the performance effectively in delay-constrained mobile ad hoc networks.

Keywords: ad-hoc networks, topology, interference, algorithm, optimization

I. INTRODUCTION

Conventionally, most of digital components which necessitate network connections in order to provide data services which in turn connected through permanent infrastructures like base stations. Various practical constrains are incorporated in communication services in the locations without predetermined infrastructures. Particularly, heterogeneous ad hoc networks consist of different types of terminal accessories, access technologies, number of receiver

(antennas), rate of transmission and power at different terminal nodes. This sort of provision could provide suppleness for wireless communication, which results in new challenges for network design and optimization. Zhang et al. (2015) attempted to compute the average end-to-end delay of CBR packets established at the target spots with increasing traffic pack. In this study, the author focused mainly on delay concern. Additionally, the transmission power in ITCD is minimized while keeping the connectivity and packet collisions are taken into account and also the mobility is also considered to remove un-stable links in the topology. ITCD can guarantee terminal destination nodes to receive data packets successfully with a large probability and make end-to-end delay within a threshold by adjusting transmission power. Li and Eryilmaz (2012) proposed an algorithm to describe the challenging problem of designing a scheduling policy for end-to-end deadline constrained traffic with reliability requirements in a multi-hop network. In their research work, the main objectives is framed orientating towards scheduling alone. Li et al. (2009) revealed that an optical network is too costly to act as a broadband access network. On the other hand, a pure wireless ad-hoc network with different nodes may not provide satisfactory broadband services since the per node throughput diminishes as the number of users increase. In this case, hybrid wireless networks have greater throughput capacity and smaller average packet delay than pure ad hoc networks. The present study proposed three different algorithms with different complexity and characteristics. The throughput capacity and the average packet delay are taken into account and the proposed protocol focuses at minimizing the overall network overhead and energy expenditure associated with the multihop data retrieval process while also ensuring balanced energy consumption among SNs and prolonged network lifetime. This is achieved through building cluster

structures consisted of member nodes that route their measured data to their assigned cluster head (CH). Also, clustering has proven to be an effective approach for organizing the network in the above context. Besides achieving energy efficiency, clustering also reduces channel contention and packet collisions, resulting in improved network throughput under high load.

II. RELATED WORKS

In wireless communications, the goal of the medium access control (MANET) protocol is to efficiently utilize the wireless medium, which is a limited resource. The effective use of the channel strongly determines the ability of the network to meet application requirements such as quality of service (QoS), energy dissipation, fairness, stability, and robustness (Rahnem, 1993). Based on the collaboration level, MANET protocols can be classified into two categories: coordinated and non-coordinated (Numanoglu et al., 2005). Channel access in non-coordinated protocols is typically based on a contention mechanism between the nodes. IEEE 802.11 (Huang and Lai, 2002) is an example of a non-coordinated protocol. Although it is easier to support non-uniform traffic with non-coordinated protocols, these protocols are unsuitable for highly loaded networks due to the contention mechanism. On the other hand, in coordinated channel access protocols, the medium access is regulated, making them better suited for networks where the network load is high. IEEE 802.15.3, IEEE 802.15.4, and MH-TRACE (Cooklev, 2004) are examples of such coordinated protocols. Coordinated channel access schemes provide support for QoS which in turn reduce energy dissipation, and increase throughput for low-to-mid noise levels and for dense networks. However, these protocols perform poorly under non-uniform traffic loads. MH-TRACE further uses a soft clustering approach where the clustering mechanism is utilized only for providing channel access to the member nodes. Hence, each node is capable of communicating directly with every other node provided that they are within communication range of each other.

The main consideration in forming clusters is the load distribution in the network. Clusters should be formed in such a way that they are able to meet the demand for channel access of the nodes in the cluster as much as possible. When the cluster is not able to meet the demand, either some of the transmissions are deferred (better suited for guaranteed delivery traffic) or the packets are dropped (better suited for best effort traffic). Thus, while designing a protocol or determining the performance of a specific protocol, the

load distribution has crucial importance. Clustering approaches may be classified as soft and hard clustering. In hard clustering approaches, such as GSM networks (Mohapatra et al., 2003), nodes belong to the cluster in which they operate.

Due to fading, two distinct transmissions may successfully operate over the same frequency, code and time range if they are well separated spatially. A successful protocol should employ this kind of spatial reuse for the sake of efficient use of the channel resources. Clustering protocols, aim to maximize the distance between the clusters using the same portion of the channel. In cellular networks, the same set of frequencies may be assigned to cells (clusters) that are separated well enough depending on the frequency reuse factor employed (Goldsmith et al., 2011). Analogously, in MH-TRACE, each cluster operates in one of several frames separated in time. MH-TRACE has internal mechanisms that maximize the distance between clusters operating in the same frame (co-frame clusters). To analyze the performance of soft clustering protocols to determine how to best set their parameters for efficient use of the channel resources. Specifically, the clustering mechanisms of MH-TRACE is described in detail as shown in the figure 1. Diamonds represent selected clusterheads (CH) and dots represent the nodes in the network. CH frame matching, together with the contents of each frame, is depicted. There are randomly chosen clusterheads that regulate the channel and provide channel access for the nodes in their communication range. Each clusterhead (CH) operates using one of the frames in the superframe structure. There is also a spatial reuse mechanism that allows more than one CH to operate in the same time frame provided that the interference is low.

Each frame in the superframe is further divided into sub-frames. The control sub-frame constitutes the management overhead. Beacon, cluster announcement(CA), and header slots of the control sub-frame are used by the CHs, whereas contention slots and information summarization (IS) slots are used by the ordinary nodes. At the beginning of the frame, the CH announces itself to the nearby nodes by sending a beacon message in the beacon slot of the control sub-frame. The CA slot is used for interference estimation for CHs operating in the same frame (co-frame CHs). During the CA slot, the CH transmits a message with a given probability and listens to the medium to calculate interference caused by other CHs operating in the same frame.

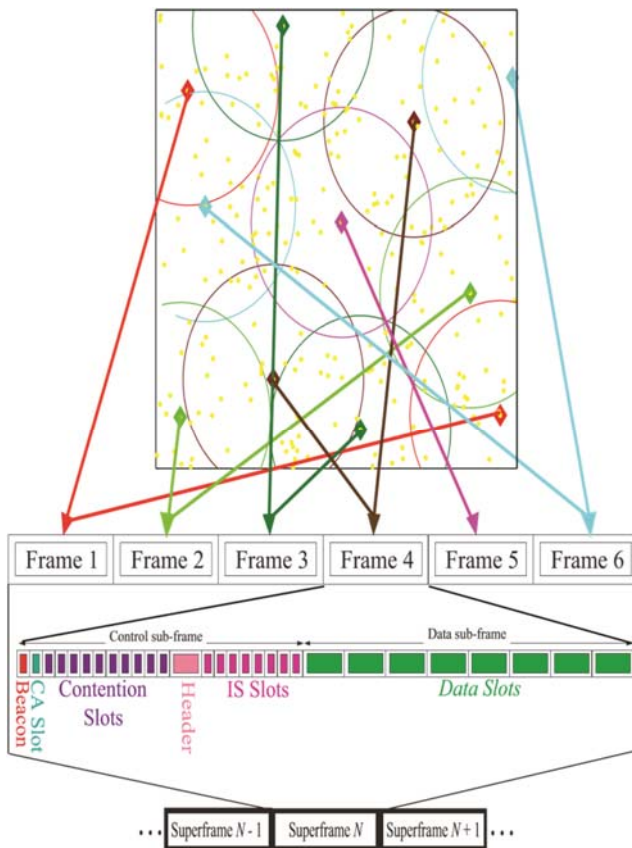


Figure 1: A snapshot of MH-TRACE clustering and medium access

Contention slots are utilized by the nodes to pass their channel access requests to the CH. A node that wants to access to the channel selects a contention slot randomly among the contention slots and sends a contention message in that slot. After listening to the medium during the contention slots, the CH becomes aware of the nodes that request channel access and forms the transmission schedule by assigning available data slots to the nodes. After that, the CH sends a header message that includes the transmission schedule that will be followed for the rest of the frame. There are an equal number of IS slots and data slots in the remainder of the frame. During the IS slots, nodes send short packets summarizing the information that they are going to be sending in the order announced in the Header. By listening to the relatively shorter IS packets, nodes become aware of the information that are going to be sent and may choose to sleep during the corresponding data slots if they are not interested in (or the recipient of) the data.

The most direct approach to determine the MANET performance is to obtain samples of field measurements on the performance metrics (Redi et al., 2006). However, the difficulty in implementation on real hardware and taking a large set of field measurements make this method impractical for most cases, and not the best approach in the protocol

design stage. It is easier and more convenient to implement a protocol on a simulation platform. Thus, simulation studies are the most widely used methods to evaluate the performance of protocols (Wang et al., 2012). However, it is impractical to determine the performance of a protocol for large sets of conditions as simulations require excessive amounts of processing power and time. Analytical models are the most suitable tools to obtain insight into the performance of a MANET protocol. Various analytical studies of protocol performance exist in the literature. These studies range from detailed protocol specific models to more general models that can be applied to a group of protocols.

III. PROBLEM FORMULATION

The present study aims to achieve efficient bandwidth and energy utilization for MANETs and specifically focuses on the MANET and the routing layers. The key challenges in effective MANET protocol design are the maximization of spatial reuse and providing support for non-uniform load distributions. Spatial reuse is tightly linked to the bandwidth efficiency. Due to the noisy nature of the propagation medium, the same channel resources can be used in spatially remote locations simultaneously without affecting each other. Incorporating spatial reuse into the MANET protocol drastically increases bandwidth efficiency. On the other hand, due to the dynamic behavior in MANETs, the traffic load may be highly non-uniform over the network area. Thus, it is crucial that the MANET protocol be able to efficiently handle spatially non-uniform traffic loads. Uncoordinated protocols intrinsically incorporate spatial reuse and adapt to the changes in load distribution through the carrier sensing mechanism. However, coordinated protocols require careful design at the MANET layer allowing the channel controllers to utilize spatial reuse and accommodate any changes in the traffic distribution.

IV. SYSTEM ARCHITECTURE

The present study adapted the following system architecture (Figure 2) to overcome the above statement of problems in effective MANET protocol design.

In the node distribution changes and packet generation patterns result in a non-uniform load distribution. Similar to cellular systems, coordinated MANET protocols need specialized spatial reuse and channel borrowing mechanisms that address the unique characteristics of MANETs in order to provide as high bandwidth efficiency as their uncoordinated counterparts.

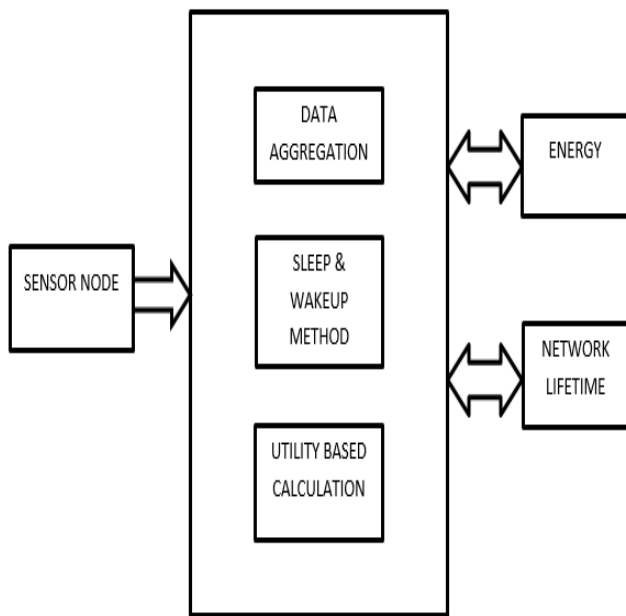


Figure 2: System Architecture

Figure 3 explains the data flow pattern in the proposed protocol design. Due to node mobility and the dynamic nature of the sources in a MANET, the network load oftentimes is not uniformly distributed. The proposed algorithm for managing non-uniform load distribution in MANETs into the MH-TRACE framework and it incorporates spatial reuse which does not provide any channel borrowing or load balancing mechanisms. Thus, it does not provide optimal support to dynamically changing conditions and non-uniform loads. Hence, intentionally the present study applies the dynamic channel allocation and cooperative load balancing algorithms to MH-TRACE, creating the new protocols of DCA-TRACE, CMH-TRACE and the combined CDCA-TRACE.

V. USER INTERFACE

In order to implement the design, the study considered the internal and external agents as actors. Figure 4 explains user case diagram which consists of actors and their relationships. The diagram represents the system/subsystem of an application. A single user case diagram captures a particular functionality of a system. The class diagram (Figure 5) is the main building block of object oriented modelling. It is used both for general conceptual modelling of the systematic of the application, and for detailed modelling translating the models into programming code. Class diagrams can also be used for data modeling.

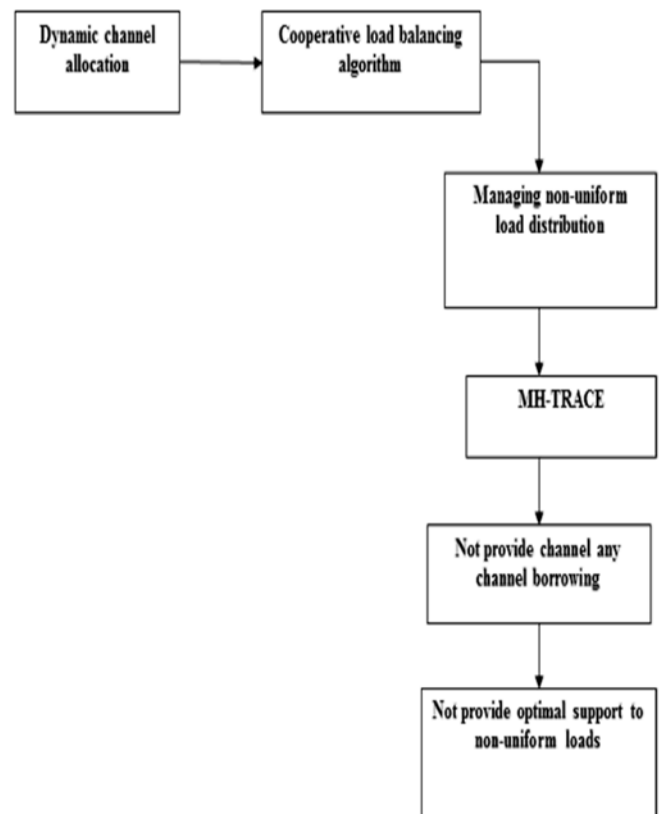


Figure 3: Data flow diagram

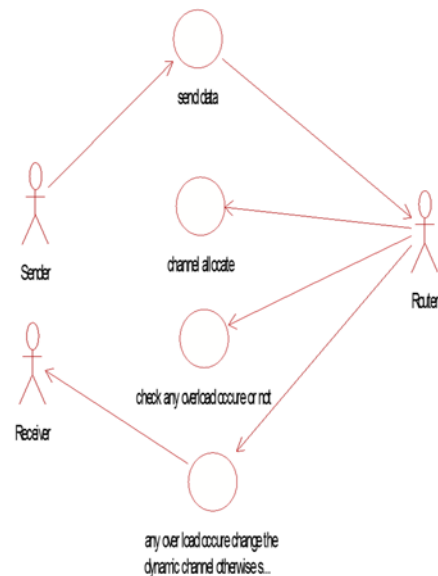


Figure 4: User Case Diagram

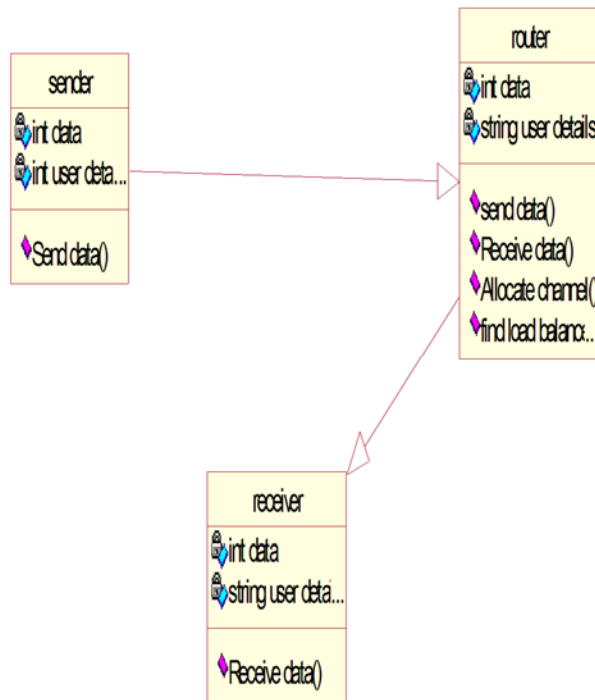


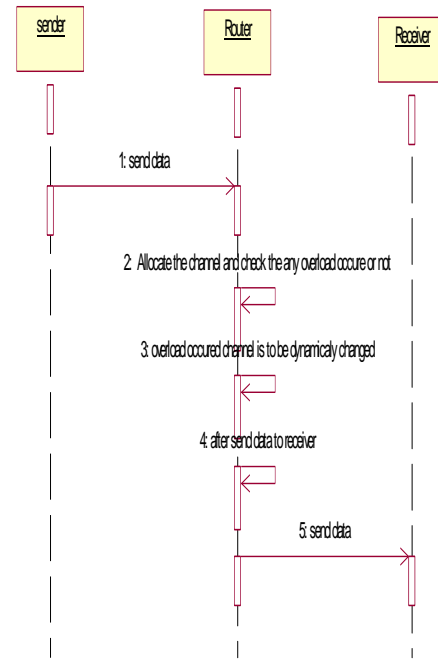
Figure 5: Class Diagram

VI. IMPLEMENTATION PHASE

The performance of reliability of the system was tested and it gained good level of acceptance. During the implementation stage a live demon was undertaken and made in front of end-users. The stage consists of the following steps.

- Testing the developed program with sample data
- Detection and correction of internal error
- Testing the system to meet the user requirement
- Feeding the real time data and retesting
- Making necessary change as described by the user

Figure 6 shows the sequence of process operates with one another and in what order. It is a construct of a message sequence chart. It also shows object interactions arranged in time sequence. It depicts the objects and classes involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario. Sequence diagrams are typically associated with use case realizations in the Logical View of the system under development. Sequence diagrams are sometimes called **event diagrams** or **event scenarios**.



VII.

Figure 6: Sequence of Process

Figure 7 resembles a flowchart that portrays the roles, functionality and behavior of individual objects as well as the overall operation of the system in real time. Objects are shown as rectangles with naming labels inside. These labels are preceded by colons and may be underlined. The relationships between the objects are shown as lines connecting the rectangles. The messages between objects are shown as arrows connecting the relevant rectangles along with labels that define the message sequencing.

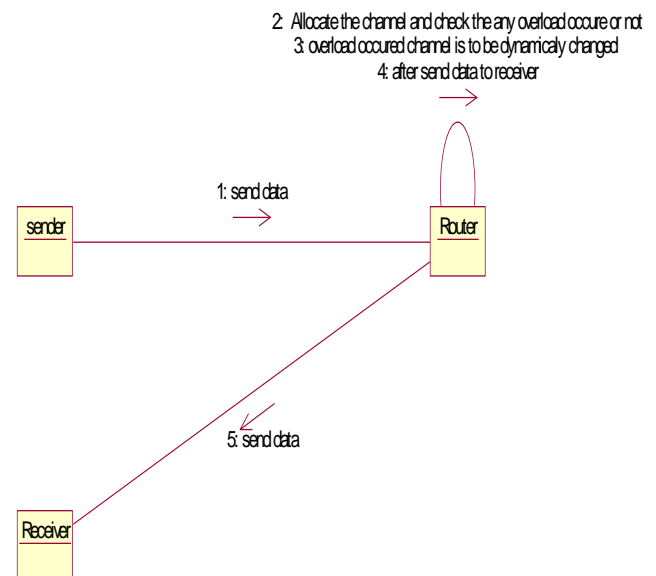


Figure 7: Collaboration Diagram

VIII.SYSTEM TESTING

As a preliminary testing, the study conducted the behavioral testing which focuses on the functional requirements of the software. It enables the software engineer to derive sets of input conditions that will fully exercise all functional requirements for a program. The study attempts to find errors in the following categories.

- Functional Testing and black box type testing geared to functional requirements of an application. This type of testing should be done by testers. Our project does the functional testing of what input given and what output should be obtained.
- System Testing-black box type testing that is based on overall requirements specifications; covers all combined parts of a system. The system testing to be done here is that to check with all the peripherals used in the project.
- Stress Testing-term often used interchangeably with 'load' and 'performance' testing. Also used to describe such tests as system functional testing while under unusually heavy loads, heavy depletion of certain actions or inputs, input of large numerical values.
- Performance Testing-term often used interchangeably with 'stresses' and 'load' testing. Ideally 'performance' testing is defined in requirements documentation or QA or Test Plans.

Additionally, the study conducted test case design method which uses the control structure of the procedural design to derive test cases.. Exercise all logical decisions on their true and false sides. Execute all loops at their boundaries and within their operational bounds. Exercise internal data structures to ensure their validity. Finally, the study implemented the most 'micro' scale of testing to test particular functions or code modules. Not always easily done unless the application has a well designed architecture with tight code; may require developing test modules or test harnesses.

IX.LIMITATIONS

The crucial challenges of implementing a MANET protocol on real hardware. The study simulation do not accurately reflect many of the challenges encountered in real implementations such as limited processing power, clock drift, synchronization, imperfect physical

layers, and cross band interference. The present research work develops a reusable hardware framework to evaluate the performance of 10 wireless protocols, in particular the TRACE protocol for real-time communication in mobile ad hoc networks. Also, the testing of TRACE implementation for packet losses and operation of the TRACE protocol depends on the cooperation and control information exchange between the nodes in the network. On the other hand, packet losses in the system disrupt the availability of such information. As an attempt, the current study adds packet loss compensation systems in the TRACE implementation to increase the robustness of the implementation against packet losses.

X. CONCLUSION

In the present study did not investigate the effects of upper layers such as the routing layer and instead focused on the MANET layer capability and local broadcasting service. The study concluded that the packet routing has a significant impact on the load distribution. Moreover, it can be used alongside with network coding and simultaneous transmission techniques for cooperative diversity. In general, joint optimization of the MANET and routing layers may enable even more efficient solutions. The investigation of the effects of routing would be considered as a future work.

REFERENCES

1. Andrews, J., Shakkottai, S., Heath, R., Jindal, N., Haenggi, M., Berry, R., et al. (2008). Rethinking information theory for mobile ad hoc networks. *IEEE Communications Magazine*, 46(12), 94–101.
2. Wang, Z., Chen, Y., & Li, C. (2012). Corman: A novel cooperative opportunistic routing scheme in mobile ad hoc networks. *IEEE Journal on Selected Areas in Communications*, 30(2), 289–296.
3. Goldsmith, A., Effros, M., Koetter, R., Medard, M., Ozdaglar, A., & Zheng, L. (2011). Beyond shannon: The quest for fundamental performance limits of wireless ad hoc networks. *IEEE Communications Magazine*, 49(5), 195–205.
4. Kannhavong, B., Nakayama, H., Nemoto, Y., Kato, N., & Jamalipour, A. (2007). A survey of routing attacks in mobile ad hoc networks. *IEEE Wireless Communications Magazine*, 14(5), 85–91.
5. Kannhavong, B., Nakayama, H., Kato, N., Jamalipour, A., & Nemoto, Y. (2007). A study of a routing attack in olsr-based mobile ad hoc networks. *International Journal of Communication Systems*, 20(11), 1245–1261.

6. Zhou, Y., Zhang, Y., Xie, Y., Zhang, H., Yang, L. T., & Min, G. (2014). TransCom: A virtual disk-based cloud computing platform for heterogeneous services. *IEEE Transactions on Network and Service Management*, 11(1), 46–59.
7. T. Numanoglu, B. Tavli, and W. Heinzelman. An analysis of coordinated and non-coordinated medium access control protocols under channel noise. *Military Communications Conference, 2005. MILCOM 2005. IEEE*, pages 2642–2648 Vol. 4, Oct. 2005.
8. A. Chandra, V. Gummalla, and J. O. Limb. Wireless medium access control protocols. *IEEE Communications Surveys and Tutorials*, 3:2–15, 2000.
9. P. Mohapatra, J. Li, and C. Gui. Qos in mobile ad hoc networks. *IEEE Wireless Communications Magazine*, 10:44–52, 2003.
10. B. Tavli and W. B. Heinzelman. MH-TRACE: Multi hop time reservation using adaptive control for energy efficiency. *IEEE Journal on Selected Areas of Communications*, 22(5):942–953, June 2004.
11. T. Cooklev. *Wireless Communication Standarts*. IEEE Press, 2004.
12. J. Karaoguz. High-rate wireless personal area networks. *Communications Magazine, IEEE*, 39(12):96–102, Dec 2001.
13. T. Numanoglu, B. Tavli, and W. B. Heinzelman. The effects of channel errors on coordinated and non-coordinated medium access control protocols. In *Proceedings of IEEE International Conference on Wireless and Mobile Computing*, volume 1, pages 58–65, Aug 2005. 261.
14. Bora Karaoglu, Tolga Numanoglu, and Wendi Heinzelman. Analytical performance of soft clustering protocols. *Ad Hoc Networks*, 9(4):635 – 651, 2011.
15. Lifei Huang and Ten-Hwang Lai. On the scalability of ieee 802.11 ad hoc networks. In *Proceedings of the 3rd ACM international symposium on Mobile ad hoc networking & computing, MobiHoc '02*, pages 173–182, New York, NY, USA, 2002. ACM.
16. IEEE 802.15.3 Working Group. Part 15.3: Wireless medium access control (MANET) and physical layer (PHY) specifications for high rate wireless personal area networks (WPAN). *IEEE Draft Standard, Draft P802.15.3/D16*, Feb 2003.
17. Mikko Kohvakka, Mauri Kuorilehto, Marko Hännikäinen, and Timo D. Hamäläinen. Performance analysis of ieee 802.15.4 and zigbee for largescale wireless sensor network applications. In *Proceedings of the 3rd ACM international workshop on Performance evaluation of wireless ad hoc, sensor and ubiquitous networks, PE-WASUN '06*, pages 48–57, New York, NY, USA, 2006. ACM.
18. M. Rahnema. Overview of the gsm system and protocol architecture. *Communications Magazine, IEEE*, 31(4):92–100, Apr 1993.
19. T. S. Rappaport. *Wireless Communications: Principles and Practice*. Prentice Hall, Upper Saddle River, NJ, USA, 2002.
20. Jason Redi, Bill Watson, Ram Ramanathan, Prithwish Basu, Fabrice Tchakountio, Michael Girone, and Martha Steenstrup. Design and implementation of a mimo MANET protocol for ad hoc networking. volume 6248, page 624802. SPIE, 2006. 262
21. V. Tippanagoudar, I. Mahgoub, and A. Badi. Implementation of the sensorMANET protocol for the jist/swans simulator. In *Computer Systems and Applications, 2007. AICCSA '07. IEEE/ACS International Conference on*, pages 225 – 232, may 2007.

Intelligent Intrusion Detection System Based on MLP, RBF and SVM Classification Algorithms: A Comparative Study

Ismail M. Keshta 

ismailk@dcc.kfupm.edu.sa

dr.ismail.keshta@gmail.com

Abstract—An effective approach for tackling network security problems is Intrusion detection systems (IDS). These kind of systems play a key role in network security as they can detect different types of attacks in networks, including DoS, U2R Probe and R2L. In addition, IDS are an increasingly key part of the system's defense. Various approaches to IDS are now being used, but are unfortunately relatively ineffective. Data mining techniques and artificial intelligence play an important role in security services. We will present a comparative study of three well-known intelligent algorithms in this paper. These are Radial Basis Functions (RBF), Multilayer Perceptrons (MLP) and Support Vector Machine (SVM). This work's main interest is to benchmark the performance of these 3 intelligent algorithms. This is done by using a dataset of about 9,000 connections, randomly chosen from KDD'99's 10% dataset. In addition, we investigate these algorithms' performance in terms of their attack classification accuracy. The Simulation results are also analyzed and the discussion is then presented. It has been observed that SVM with a linear kernel (Linear-SVM) gives a better performance than MLP and RBF in terms of its detection accuracy and processing speed.

Keywords- Intrusion detection system; Network security; Machine learning; Anomaly detection; KDD Cup 99

I. INTRODUCTION

Network security is fast becoming a big challenge. As interconnections among computer systems grow rapidly Computer networks need to be protected against the unauthorized disclosure of information, denial-of-service (DoS) attacks and the modifying or destroying of data [1].

Attack detection techniques have become a critical issue that are being used to secure networks. Making a network secure is so difficult for many reasons, including the complexity of computers and networks, a lack of awareness of the various risks and threats, increasing internet usage and the computer system's vulnerabilities [2][3]. It is vital to note here that detection techniques have become a vital difficulty of open research and so they get given the additional attention of the research community. Furthermore, it is important to state that the network attacks' complex properties are key issues that work against these detection techniques [4][5].

The traditional techniques, including avoiding any programming errors and firewalls, have not succeeded in fully protecting networks and systems from the dangers of malware

and so attacks are becoming increasingly sophisticated [6]. Peddabachigari *et al.* [7] showed that programming errors can no longer be avoided as the system's complexity and application software is rapidly evolving, leaving weaknesses that can be exploited. Jamali *et al.* [8] state that firewalls are not sufficient to give the network total security because they just throttle attacks that come from outside and do not have any effect on the risk of inside attacks. It is likely that computer systems will remain unsecured in the near future.

Therefore, IDS have now become a vital and indispensable part of security infrastructure that are used to detect any sophisticated attacks and malware early before they can inflict any wide spread damage [7]-[9]. IDS is, therefore, needed as an extra wall to protect systems despite these prevention techniques. Detection of intrusion is useful in the detection of intrusions that are successful, as well as monitoring bids to break security [10]-[12]. IDS protects computer systems against hateful operations by detecting the violation of security policies and active defenders, including by alarming operators [13]. It particularly helps the network to provide resistance against external attacks [14].

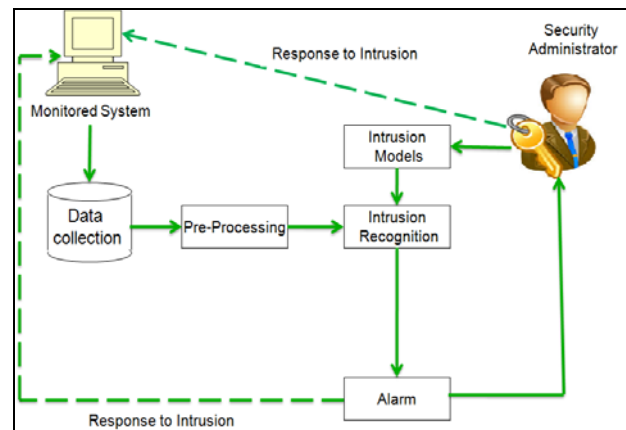


Figure 1: Organization of a generalized IDS

It is vital to state that many issues need to be considered when building an IDS, including data collection, response, data preprocessing, reporting and intrusion recognition, which is at the heart of it. The organization of an IDS is illustrated in Figure 1.

Existing IDS systems are able to be divided into two categories in general, according to each of the detection approaches, which are anomaly detection and misuse detection [15] [16]. Misuse-based IDS is able to detect known attacks efficiently, but fails to find new attacks which fail to embody the rules in the database [17]. Therefore, a database has to be continuously updated to store the signatures of every attack that is known. This IDS type is obviously unable to detect new attacks unless it is trained to [18]. Anomaly-based IDS can build a normal behavior model and it then distinguishes any major deviations from the model as being an intrusion. This IDS type is able to detect new attacks or unknown ones but it features a high rate of false alarms [19]-[20].

Research efforts have been made to reduce these false alarms by proposing an intelligent IDS that is based on machine learning. A number of anomaly detection systems are developed in literature that are based on a lot of different kinds of machine learning techniques[21]-[25]. Some of these studies can apply single learning techniques. But some systems are based on a combination of different learning techniques. Machine learning classification algorithms provide a very promising solution and are able to discover novel attacks that are based on their own features[26]. In addition, they can be utilized to study and then identify correlated data, make decisions, make predictions and classify data[24]-[26].

Algorithms like Multilayer Perceptron (MLP), Radial Basis Functions (RBF) and Support Vector Machine (SVM) are all examples of algorithms which are well-known, widely adopted and have been investigated in neural networks, machine learning and artificial intelligence. MLP can, for a start, successfully perform the classification operation[27][28], while MLP neural network training is hard because of its structure's complexity [28]. SVM is also a very strong algorithm in data mining, which has been applied successfully in a number of scientific applications [29].

Despite how vital machine learning algorithms are for intrusion detection systems[21]-[25], more could be done to provide comparison studies between the algorithms, as little attention has been given to this, particularly when it comes to the designing of an effective IDS for both computer and network systems. Furthermore, little has been done to specify an intelligent IDS that would reduce the anomaly-based detection's false alarm rate.

We conducted in this work a comprehensive and detailed comparative study across a total of 3 intelligent classification algorithms, which are RBF, SVM and MLP, with linear kernel. This is a polynomial kernel with exponent 1, and we chose it to be linear for SVM as Linear-SVM is both efficient and fast. Linear-SVM is able to consume less energy in the course of the learning process in the deployment phase, unlike MLP and RBF [30] [31]. Moreover Gupta and Ramanathan [32], as well as Magno *et al.* [33], stated that Linear-SVM is a low complexity classifier. Magno *et al.* [33] also highlighted that Linear-SVM gives a good balance between the computational and memory cost and the percentage of correctly classified data. Sazonov *et al.* [34] said there were two powerful SVM characteristics, which are high generalization and robustness. In addition, Bal *et al.* [35] found that SVM with linear kernel is

a very promising algorithm that exists in the machine learning field. Yuan *et al.* [36] also concluded that an SVM classifier, especially one with linear kernel, can both learn and build the knowledge that is needed from less training samples and yet can still provide a high level of classification accuracy, unlike a number of other classifiers such as MLP and RBF.

The following are our major contributions in this work. Firstly, we provided a number of detailed and state-of-the-art related IDS models, which were based on the intelligent machine learning algorithms. Secondly, we undertook a comprehensive comparison between three intelligent classifiers by using a real benchmark dataset. Thirdly, the performance of all three was examined by utilizing confusion matrix. Lastly, we were able to propose an intelligent IDS framework for effective and efficient IDS management computer and network systems. The framework was addressed at classification level. Utilizing Linear-SVM as an intelligent classifier, it is considered a core element in the building of the framework. We also discussed an evaluation of the proposed framework, and the simulation of results for detecting malicious attacks, like Remote to Local (R2L), Denial of Service (DoS) Attacks, Remote to User (R2U) Attacks and Probing attacks, are all provided.

The rest of this paper is organized in the following way. Section II gives a literature review of the recent approaches that have been proposed for IDS that is based on intelligent classification algorithms. Section III highlights some background into the classification algorithms that were utilized in the work, which is RBF, MLP and SVM. It also provides a useful overview of the experimental dataset. The paper's main contribution is discussed in Section IV, while simulation experiments and the ways they were setup is illustrated in Section V. This section also summarizes and discusses the results of the simulation. Lastly, Section VI provides the conclusion of the paper and also highlights any future research directions.

II. RELATED WORK

There has been a lot of researches into anomaly-based intrusion detection, and some of them have used machine learning, as well as data mining techniques. Decision tree, neural networks, clustering and Bayesian parameter estimation are some techniques that have been used to detect any intrusive behaviors in the computer network.

Chandollikar *et al.* [37] evaluated the performance of 2 classification algorithms. These were Bayes net and J48 algorithm, which are both used for detecting computer attacks. The results reveal that J48 learning algorithm was more accurate than Bayes net algorithm in terms of achieving better accuracy and it had a lower error rate. A benchmark was used in the evaluation. This was the KDD cup dataset. It was emphasized that J48 algorithm had a higher accuracy which helps to increase the IDS' efficiency.

The Principal Component Analysis and Naive Bayes classifier was employed by Panda *et al.* [38] to give them a way of detecting intrusion by using machine learning algorithms. These experiments were carried out on the KDD'99 cup dataset, an intrusion detection dataset. The dimensionality

of the dataset was reduced by utilizing principal component analysis, as well as the Naïve Bayes classifier classification of the dataset. This was done in both the normal and attack classes. They concluded that the approach they used was a description of a Network Intrusion detection system framework which used two algorithms, Naive Bayes and Principal Component Analysis. The result they obtained showed that their approach was faster compared to a number of the other existing systems.

An intrusion detection system was proposed by Wang *et al.* [39]. This was based on C4.5 decision tree, one of the algorithm-based neural networks. The result revealed that the intrusion detection system was effective and feasible, and had a high rate of accuracy. All of their experiments were conducted on a KDD CUP1999 dataset, a test set that is widely used for intrusion detection fields. The tree that is generated by the C4.5 neural network classification algorithm for intrusion detection was used to build rules. These then can use the knowledge base of IDS. In other words, the rules are able to give an indication if a new network behavior is either normal or abnormal, based on the built knowledge.

The J48 intelligent algorithm was utilized by Chandolika *et al.* [40] in the experiments they did to make IDS. Their results show that J48 is an effective and efficient algorithm of the classification in the KDD CUP1999 dataset.

Yogita *et al.* [41] proposed IDS that used SVM as a data mining technique. It is vital to mention here that SVM is a very popular classification algorithm. However, they highlighted the main drawback, which is that SVM takes a very long time to train the neural network. These experiments were done by utilizing the NSL-KDD Cup'99 dataset's improved version of the KDD Cup'99 dataset. They used the Gaussian RBF as the kernel function and a 10-fold cross validation as the test option parameter that was used for SVM. In addition, they pointed out that the method based SVM that was proposed was able to increase the accuracy of intrusion detection and cut down on the time taken to build this classification model.

The aim of Mohammadreza *et al.* [42] was to use data mining techniques, which included SVM and the classification tree for IDS. The results reveal that the C4.5 algorithm is better than SVM at detection of any network intrusions. These experiments were carried out on a KDD CUP 99 dataset. Das *et al.* [43] looked at the IDS at its preprocessing level, which is the level before the classification process, and proposed what is called a divide and conquer algorithm. The aim of this was to reduce the feature set from the large KDD 99 dataset. The proposed algorithm successfully reduced the IDS's overhead for analyzing the entire KDD dataset. This was done by selecting the vital features and then classifying them all with a maximized rate of classification. It was a generic algorithm and it could be applied to absolutely any dataset. The authors used LDA, KNN, C4.5, SVM and a number of classification algorithms in order to classify the various feature sets that had been obtained.

III. PRELIMINARIES

This section gives a brief background about the three intelligent algorithms used in this study, as well as about the dataset for the experimental comparison.

A. Classification Algorithms

The various classification algorithms that were used in the research project are described in brief below.

1) Multilayer Perceptron (MLP)

This is composed of a big amount of widely interconnected neurons that all work in parallel in order to solve a particular problem. MLP is organized in a series of layers that have a feed-forward information flow. An MLP network's main architecture consists of a number of signals which flow sequentially through these various layers, starting with the input layer, through to the output layer. Between these two layers are a number of intermediate layers, which are also known as hidden layers because you cannot see them at either the input or the output. Each of the units is first utilized to calculate what the difference is between a vector of weights and a vector provided by the outputs of the previous layer. In order to generate the next layer's input, a transfer function, which is also called activation, was applied to the result [44]. RBF, unipolar sigmoid and bipolar sigmoid are all examples of activation functions that are both well-known and commonly used.[45]. The training phase's main steps in an MLP network are the following: Firstly, after being given the dataset's input pattern, this particular pattern is forward-propagated to the MLP network's output and it is then compared with the output desired. Secondly, the error signal that exists between the network's output and the desired response is then back-propagated to the network. Lastly, a number of adjustments are made to the synaptic weights [46]. The process is repeated for the next input vector and this continues until all of the training patterns have been passed right through the network.

2) Radial Basis Functions (RBF)

This involves a total of three layers. The first is called the input layer and it is made up of source nodes (or sensory units). The amount of these source nodes is equal to the input vector's dimension. The second is the hidden layer, which consists of nonlinear units. These are directly connected to every one of the sensory units in the input layer. The RBF network has only a single hidden layer that has RBF activation functions. Lastly, the output layer is utilized to linearly combine the hidden layer's outputs and give the network's response to the input data [47].

3) Support Vector Machine (SVM)

This splits the dataset into two different classes. These are separated by placing a linear boundary between both the normal and attack classes in a way that maximizes the margin. SVM finds the hyperplane that is able to provide the maximum distance there is between the hyperplane and the closest of the positive and negative samples [48][49]. The SVM network's basic structure is similar to the structure of the ordinary RBF network. However, the kernel activating function is applied instead of the exponential activating function (which is generally Gaussian activation functions). This kernel activating function can be either a polynomial kernel, a Gaussian radial

basis kernel, or two layer feed-forward neural network kernels [49].

B. Dataset

This section gives a brief description of the dataset that is used in the work. The KDDCUP'99 dataset was prepared by the 1998 DARPA Intrusion Detection Evaluation program by MIT Lincoln Laboratories [50]. It can be seen from the literature that this dataset has been used widely for the evaluation of anomaly based IDS. A lot of researchers are using the KDDCUP'99 dataset as it is the only publicly available dataset for the ID problem, and also because it is possible to extract useful information from it [51] [52]. The full dataset contained around 5 million instance/records. This is where each data row has its connection records. Connection is defined in many references as a sequence of TCP packets that start and end at some time between a source and a destination under a protocol that is well-defined [50]-[52].

This dataset contains a number of different attack types, which are classified into 4 major categories. These are R2L, DOS, Probing and U2R. The KDD cup 99 set has a total of 41 attributes or features for each instance, or a sample plus 1 class label. The total number is, therefore, 42 attributes. The 41 attributes are destination bytes, count, dst host count, diff srv rate, wrongfragment and urgent. The 42nd field is a label that can be generalized as either normal or anomaly (U2R, DoS, Probing and R2L) [50] [53] (see Table 1).

TABLE 1: TYPES OF ATTACKS IN KDD'99 DATASET

| Classification | Short Description | Name of Attacks |
|----------------|--|---|
| DoS | Attacker attempts to deny or prevent legitimate users from using a service. | smurf, land, pod, teardrop, neptune, back |
| R2L | Attacker attempts to send packets to the victim machine in order to gain access because he does not have an account on it. | ftp_write, phf, spy, warezmaster, warezclient, imap, guess_passwd, multihop |
| U2R | The attacker tries to exploit some vulnerability to gain root/super user access to the system. | perl, buffer_overflow, rootkit, loadmodule |
| Probe | The attacker attempts to gather information about a computer network. | portsweep, nmap, ipsweep, satan |

Tavallae *et al.* [26] highlighted that the features of KDD'99 can be categorized into three different groups. These are Basic features, Content features and Traffic features. Basic features are utilized to encapsulate all of the attributes that have been extracted from the TCP/IP connection.

The majority of these features can help to detect the major causes of network delays. There is then a second class, which is the Traffic features. These depend on window interval and they can be divided into 2 major features, which are "same host" features and also "same service" features. They are, therefore, called time-based features. "Same host" features are used to carry out an examination of network connections in the previous two seconds, and they have the same target host as the

current connection. "Same service" features are utilized to test the network's connections and have the same service as the current connection in the previous two seconds. The last of these classes is called Content features, which helps to detect U2R and R2L attacks. This is because these types of attacks do not have either a well-defined structured feature or well-defined pattern. Therefore, Content features have some features that enable IDS to detect any intrusion that is tending to cause or create suspicion in the data portion, like a number of failed log on attempts [26].

IV. THE PROPOSED SYSTEM

The focus of this research work is on the original "10% KDD 99" dataset because of the limited memory capacity. The system flow for the proposed IDS is shown in Figure 2. The original "10% KDD 99" dataset is firstly loaded into the system. The next step is pre-processing, in which the input file is properly prepared.

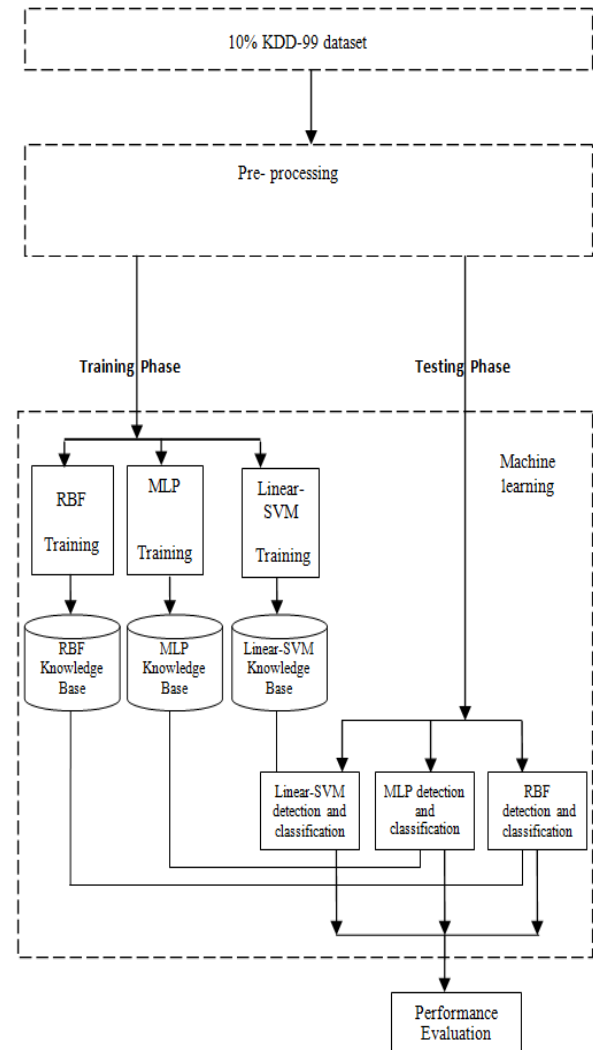


Figure 2: Block diagram of the proposed IDS

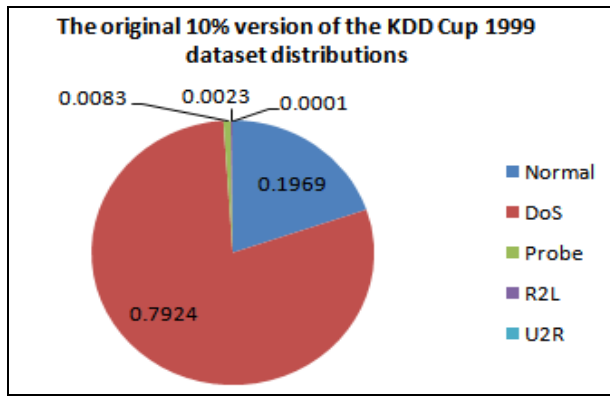


Figure 3: Original KDD'99 10% dataset distributions

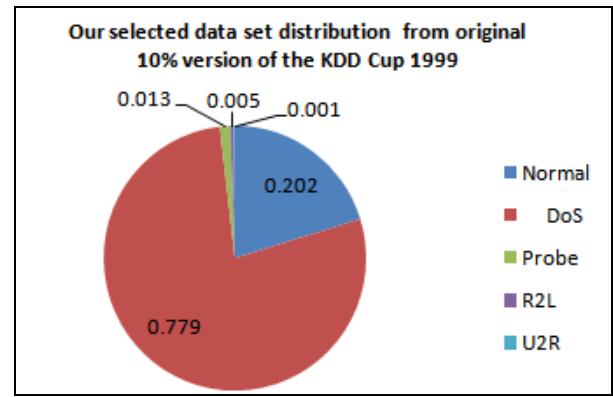


Figure 4: Selected dataset distributions

In this step, a total of 9000 instances are randomly selected from 10% of the KDD CUP 1999 dataset with nearly the same distribution as the KDD'99 10% dataset. Figure 3 and Figure 4 clearly highlighted this point. Two phases are performed after that. Firstly, the training/learning phase, which enables the intelligent system to build up the right knowledge base. IDS learns about relationships that exist in the built training dataset. This training phase is seen as an adaptation process to IDS in order to give the best response during the next phase, which is the testing phase.

In this phase, the intelligent system will receive different dataset, testing data set and processes it to produce an output. To test and evaluate MLP RBF and Linear-SVM algorithms, a 5-fold cross validation is utilized as a test option. The dataset is split into 5 subsets, and for each running time, one of these five subsets are used as the training set and then the other subsets as the test set.

In order to evaluate the algorithms' effectiveness for IDS, three experiments are carried out. The WEKA simulator version 3.6 [54] is utilized in the classification process. That is, the available algorithms for RBF, MLP and SVM on the Weka simulator are employed. For the Weka parameters of the algorithms, the Weka system's default settings are utilized, except for the fold cross validation, where we utilized value five.

V. DISCUSSION OF RESULTS

The confusion matrix is used to measure the three intelligent algorithms' performance [55][56]. This provides visualization of how the classifier performs on the input dataset. A number of different performance metrics, including recall, accuracy and specificity, are derived from the confusion matrix. Table 2 shows the structure of this matrix. The 4 possible outcomes/cases are true positive (TP), false positive (FP), and false negative (FN) and true negative (TN) [51][57].

TABLE 2: CONFUSION MATRIX

| | | Predicted class | |
|--------------|----------|-----------------|----------|
| | | Positive | Negative |
| Actual class | Positive | TP | FP |
| | Negative | FN | TN |

We evaluated these algorithms by using accuracy as the performance metric in this study. Accuracy in this instance represents the overall correctness of the intelligent classification of the dataset. It is given by:

$$\text{Accuracy} = \frac{(TN + TP)}{(TN + TP + FP + FN)}$$

As shown in Figure 5, the obtained results out from our dataset show the comparison between the three intrusion detection systems.

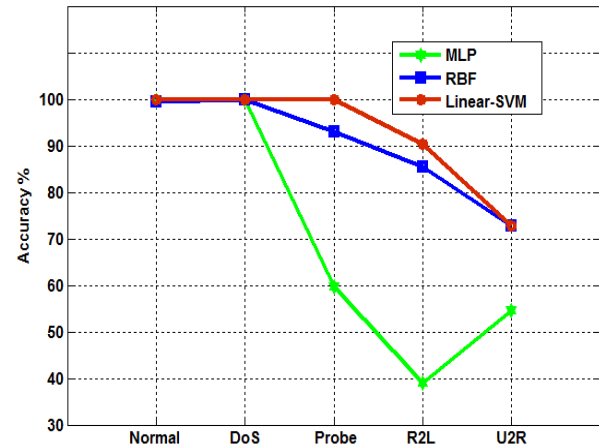


Figure 5: Accuracy comparison graph between MLP, RBF and Linear-SVM as classifiers and cross Validation (folds-5) as Test Option over our selected dataset

If we compare RBF, MLP and SVM (linear kernel), we can see that under the cross Validation Method (5-fold) Test Option, it is SVM with linear kernel that has the highest identification of correct instances (it is 99.84 % $((1817+7010+114+37+8)/9000 \times 100 = 99.84\%)$). The second highest is RBF, which is around 99.64%.

MLP has the least with 98.98%. It is worth noting that when it comes to the average time to build the model, RBF proves to be much faster than MLP as the hidden layer is computed through a single function, rather than a series of weights, as is the case with MLP. It therefore can be concluded that Linear-SVM provides the highest accuracy and the lowest error rates. We can therefore generally conclude that the SVM's performance with linear kernel was the best of the other

classifiers in detecting these attacks. So it is more accurate than either RBF or MLP. In addition, Linear-SVM is the quickest classifier in terms of building the detection model, compared to either RBF and MLP.

Figure 6 and Figure 7, respectively, show the accuracy and error rates for the three algorithms. It is also important to note that Tables 3, 4, and 5 highlight the Confusion matrixes for RBF, MLP and Linear-SVM, respectively.

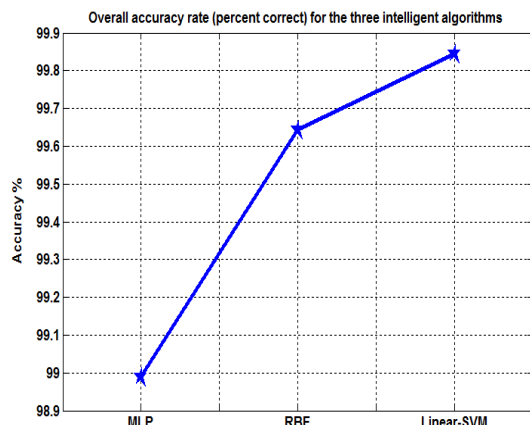


Figure 6: Overall accuracy rate for the three intelligent algorithms

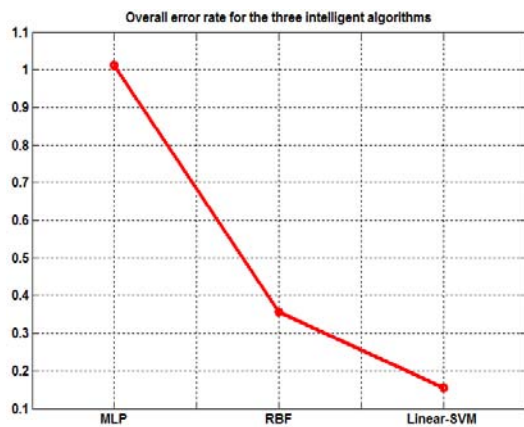


Figure 7: Overall error rate for the three intelligent algorithms

TABLE 3: CONFUSION MATRIX FOR MLP AS CLASSIFIER OVER OUR SELECTED DATASET

| | Normal | DoS | Probe | R2L | U2R | Accuracy % |
|--------|--------|------|-------|-----|-----|------------|
| Normal | 1815 | 4 | 1 | 0 | 1 | 99.67 |
| DoS | 7 | 7004 | 2 | 0 | 0 | 99.87 |
| Probe | 24 | 22 | 68 | 0 | 0 | 59.65 |
| R2L | 9 | 7 | 9 | 16 | 0 | 39.02 |
| U2R | 1 | 4 | 0 | 0 | 6 | 54.55 |

The confusion matrixes show the number of instances that have been assigned to each class. They show how many instances for each class received various classifications. The sum of the diagonals represents the amount of samples that are correctly classified. For example, the total amount of samples for MLP that have been correctly classified is the sum of 1815, 7004, 68, 16 and 6.

TABLE 4: CONFUSION MATRIX FOR RBF NETWORK AS CLASSIFIER OVER OUR SELECTED DATASET

| | Normal | DoS | Probe | R2L | U2R | Accuracy % |
|--------|--------|------|-------|-----|-----|------------|
| Normal | 1809 | 1 | 5 | 5 | 1 | 99.34 |
| DoS | 3 | 7010 | 0 | 0 | 0 | 99.957 |
| Probe | 8 | 0 | 106 | 0 | 0 | 92.98 |
| R2L | 3 | 0 | 0 | 35 | 3 | 85.37 |
| U2R | 2 | 0 | 0 | 1 | 8 | 72.73 |

TABLE 5: CONFUSION MATRIX FOR LINEAR-SVM AS CLASSIFIER OVER OUR SELECTED DATASET

| | Normal | DoS | Probe | R2L | U2R | Accuracy % |
|--------|--------|------|-------|-----|-----|------------|
| Normal | 1817 | 3 | 0 | 0 | 1 | 99.78 |
| DoS | 3 | 7010 | 0 | 0 | 0 | 99.96 |
| Probe | 0 | 0 | 114 | 0 | 0 | 100 |
| R2L | 3 | 0 | 0 | 37 | 1 | 90.24 |
| U2R | 2 | 0 | 0 | 1 | 8 | 72.73 |

VI. CONCLUSIONS AND FUTURE WORK

Network intrusion detection has recently become an area of rapid advancement. There are similar advances in intelligent computing, which have led to several classification techniques being introduced to identify network traffic and differentiate it into anomalous and normal. Intrusion detection that is based on computational intelligence has been attracting much interest from researchers in the research community. Its characteristics, including adaptation, high computational speed, fault tolerance, and error resilience in the face of noisy information, fit the requirements that are needed to build a good intrusion detection system.

In this paper, we have explained the requirement to apply intelligent algorithms to network events in order to classify network attack events. In particular, the performance of the 3 intelligent algorithms, which are MLP, RBF and Linear-SVM, on an adapted KDD 1999 dataset was evaluated. This was by done by both simulation and a comparison study. The results obtained reveal that SVM with linear kernel will perform better than MLP and the RBF network for detecting attacks in terms of achieving better accuracy and a lower error rate.

Experiments show that Linear-SVM proves to be an efficient algorithm that is able to detect various kinds of intrusions/attacks in network, such as DoS, Probe, U2R and R2L. Linear-SVM has the best detection accuracy when it comes to detecting different types of attacks. It, therefore, has the lowest error rate of all.

As future work, we intend to evaluate SVM's under the other benchmarking datasets. In addition, we will conduct a performance comparison between SVM and different kernels, such as Gaussian or sigmoid kernels. This will be done to find the best kernel or activation function for SVM which can give the best attack detection rate for building IDS.

REFERENCES

- [1] Depren, O., Topallar, M., Anarim, E., & Ciliz, M. K. (2005). An intelligent intrusion detection system (IDS) for anomaly and misuse detection in computer networks. *Expert systems with Applications*, 29(4), 713-722.

- [2] MeeraGandhi, G. (2010). Machine learning approach for attack prediction and classification using supervised learning algorithms. *Int. J. Comput. Sci. Commun.*, 1(2).
- [3] Nguyen, H. A., & Choi, D. (2008). Application of data mining to network intrusion detection: classifier selection model. In *Asia-Pacific Network Operations and Management Symposium* (pp. 399-408). Springer Berlin Heidelberg.
- [4] Subramanian, S., Srinivasan, V. B., & Ramasa, C. (2012). Study on classification algorithms for network intrusion systems. *Journal of Communication and Computer*, 9(11), 1242-1246.
- [5] Li, M., & Dongliang, W. (2009). Anomaly intrusion detection based on SOM. In *Information Engineering, 2009. ICIE'09. WASE International Conference on* (Vol. 1, pp. 40-43). IEEE.
- [6] Summers, R. C. (1997). *Secure computing: threats and safeguards*. McGraw-Hill, Inc..
- [7] Peddabachigari, S., Abraham, A., Grosan, C., & Thomas, J. (2007). Modeling intrusion detection system using hybrid intelligent systems. *Journal of network and computer applications*, 30(1), 114-132.
- [8] Jamali, S., & Jafarzadeh, P. (2011). An intelligent intrusion detection system by using hierarchically structured learning automata. *Neural Computing and Applications*, 1-8.
- [9] Wu, S. X., & Banzhaf, W. (2010). The use of computational intelligence in intrusion detection systems: A review. *Applied Soft Computing*, 10(1), 1-35.
- [10] Sundaram, A. (1996). An introduction to intrusion detection. *Crossroads*, 2(4), 3-7.
- [11] Chimedtsen, E., Iwai, K., Tanaka, H., & Kurokawa, T. (2014, December). Intrusion detection system using Discrete Fourier Transform. In *Computational Intelligence for Security and Defense Applications (CISDA), 2014 Seventh IEEE Symposium on* (pp. 1-5). IEEE.
- [12] Igbe, O., Darwish, I., & Saadawi, T. (2016). Distributed Network Intrusion Detection Systems: An Artificial Immune System Approach. In *Connected Health: Applications, Systems and Engineering Technologies (CHASE), 2016 IEEE First International Conference on* (pp. 101-106). IEEE.
- [13] Di Pietro, R., & Mancini, L. V. (Eds.). (2008). *Intrusion detection systems* (Vol. 38). Springer Science & Business Media.
- [14] Kulothungan, K., Ganapathy, S., Yogesh, P., & Kannan, A. An Agent based Intrusion Detection System for Wireless Sensor Networks Using Multilevel Classification. *International Journal of Modern Engineering Research (IJMER)*, 1(2), 55-60.
- [15] Anderson, J. A. (1995). *An introduction to Neural Networks*, MIT Press.
- [16] Rhodes, B. C., Mahaffey, J. A., & Cannady, J. D. (2000). Multiple self-organizing maps for intrusion detection. In *Proceedings of the 23rd national information systems security conference* (pp. 16-19).
- [17] Al-Yaseen, W. L., Othman, Z. A., & Nazri, M. Z. A. (2017). Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system. *Expert Systems with Applications*, 67, 296-303.
- [18] Chen, C. M., Chen, Y. L., & Lin, H. C. (2010). An efficient network intrusion detection. *Computer Communications*, 33(4), 477-484.
- [19] Deepa, A. J., & Kavitha, V. (2012). A comprehensive survey on approaches to intrusion detection system. *Procedia Engineering*, 38, 2063-2069.
- [20] Thaseen, S., & Kumar, C. A. (2013). An analysis of supervised tree based classifiers for intrusion detection system. In *Pattern Recognition, Informatics and Mobile Engineering (PRIME), 2013 International Conference on* (pp. 294-299). IEEE.
- [21] Feng, W., Zhang, Q., Hu, G., & Huang, J. X. (2014). Mining network data for intrusion detection through combining SVMs with ant colony networks. *Future Generation Computer Systems*, 37, 127-140.
- [22] Kuang, F., Xu, W., & Zhang, S. (2014). A novel hybrid KPCA and SVM with GA model for intrusion detection. *Applied Soft Computing*, 18, 178-184.
- [23] Horng, S. J., Su, M. Y., Chen, Y. H., Kao, T. W., Chen, R. J., Lai, J. L., & Perkasa, C. D. (2011). A novel intrusion detection system based on hierarchical clustering and support vector machines. *Expert systems with Applications*, 38(1), 306-313.
- [24] Hasan, M., Nasser, M., Pal, B., & Ahmad, S. (2013). Intrusion detection using combination of various kernels based support vector machine. *International Journal of Scientific & Engineering Research*, 4(9), 1454-1463.
- [25] Mukkamala, S., Sung, A. H., & Abraham, A. (2003). Intrusion detection using ensemble of soft computing paradigms. In *Intelligent systems design and applications* (pp. 239-248). Springer Berlin Heidelberg.
- [26] Tavallae, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). A detailed analysis of the KDD CUP 99 data set. In *Computational Intelligence for Security and Defense Applications, 2009. CISDA 2009. IEEE Symposium on* (pp. 1-6). IEEE.
- [27] Hassim, Y. M. M., & Ghazali, R. (2012). Training a functional link neural network using an artificial bee colony for solving a classification problems. *arXiv preprint arXiv:1212.6922*.
- [28] Pal, A. K., & Pal, S. (2013). Classification model of prediction for placement of students. *International Journal of Modern Education and Computer Science*, 5(11), 49.
- [29] Purnami, S. W., Zain, J. M., & Heriawan, T. (2011). An alternative algorithm for classification large categorical dataset: k-mode clustering reduced support vector machine. *International Journal of Database Theory and Application*, 4(1), 19-30.
- [30] Barnawi, A. Y., & Keshta, I. M. (2014). Energy management of wireless sensor networks based on multi-layer perceptrons. In *European Wireless 2014; 20th European Wireless Conference; Proceedings of* (pp. 1-6). VDE.
- [31] Barnawi, A. Y., & Keshta, I. M. (2016). Energy Management in Wireless Sensor Networks Based on Naive Bayes, MLP, and SVM Classifications: A Comparative Study. *Journal of Sensors*, 2016.
- [32] Gupta, G. R., & Ramanathan, P. (2007). Level set estimation using uncoordinated mobile sensors. In *International Conference on Ad-Hoc Networks and Wireless* (pp. 101-114). Springer Berlin Heidelberg.
- [33] Magno, M., Brunelli, D., Zappi, P., & Benini, L. (2010). Energy efficient cooperative multimodal ambient monitoring. In *European Conference on Smart Sensing and Context* (pp. 56-70). Springer Berlin Heidelberg.
- [34] Sazonov, E. S., & Fontana, J. M. (2012). A sensor system for automatic detection of food intake through non-invasive monitoring of chewing. *IEEE sensors journal*, 12(5), 1340-1348.
- [35] Bal, M., Amasyali, M. F., Sever, H., Kose, G., & Demirhan, A. (2014). Performance evaluation of the machine learning algorithms used in inference mechanism of a medical decision support system. *The Scientific World Journal*, 2014.
- [36] Yuan, S., Liang, D., Qiu, L., & Liu, M. (2012). Mobile multi-agent evaluation method for wireless sensor networks-based large-scale structural health monitoring. *International Journal of Distributed Sensor Networks*.
- [37] Chandollikar, N. S., & Nandavadekar, V. D. (2012). Comparative Analysis of Two Algorithms for Intrusion Attack Classification Using KDD CUP Dataset. *International Journal of Computer Science and Engineering (IJCSE)*, 1(1), 81-88..
- [38] Panda, M., & Patra, M. R. (2007). Network intrusion detection using naive bayes. *International journal of computer science and network security(IJCSNS)*, 7(12), 258-263.
- [39] Wang, J., Yang, Q., & Ren, D. (2009). An intrusion detection algorithm based on decision tree technology. In *Information Processing, 2009. APCIP 2009. Asia-Pacific Conference on* (Vol. 2, pp. 333-335). IEEE.
- [40] Chandollikar, N. S., & Nandavadekar, V. D. (2012). Efficient algorithm for intrusion attack classification by analyzing KDD Cup 99. In *Wireless and Optical Communications Networks (WOCN), 2012 Ninth International Conference on* (pp. 1-5). IEEE.
- [41] Bhavsar, Y. B., & Waghmare, K. C. (2013). Intrusion detection system using data mining technique: Support vector machine. *International Journal of Emerging Technology and Advanced Engineering*, 3(3), 581-586.
- [42] Ektefa, M., Memar, S., Sidi, F., & Affendey, L. S. (2010). Intrusion detection using data mining techniques. In *Information Retrieval &*

- Knowledge Management, (CAMP), 2010 International Conference on* (pp. 200-203). IEEE.
- [43] Das, A., & Nayak, R. B. (2012). A divide and conquer feature reduction and feature selection algorithm in KDD intrusion detection dataset. In *Sustainable Energy and Intelligent Systems (SEISCON 2012), IET Chennai 3rd International on* (pp. 1-4). IET.
 - [44] Battiti, R., Brunato, M., & Mascia, F. (2008). *Reactive search and intelligent optimization* (Vol. 45). Springer Science & Business Media.
 - [45] Karlik, B., & Olgac, A. V. (2011). Performance analysis of various activation functions in generalized MLP architectures of neural networks. *International Journal of Artificial Intelligence and Expert Systems*, 1(4), 111-122.
 - [46] Bouzgou, H., & Benoudjit, N. (2011). Multiple architecture system for wind speed prediction. *Applied Energy*, 88(7), 2463-2471.
 - [47] Haykin, S. S. (2001). *Neural networks: a comprehensive foundation*. Tsinghua University Press.
 - [48] Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media.
 - [49] Bennett, K. P., & Campbell, C. (2000). Support vector machines: hype or hallelujah?. *ACM SIGKDD Explorations Newsletter*, 2(2), 1-13.
 - [50] KDD Cup 1999 Data, Information and Computer Science, University of California, Irvine. <http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html>
 - [51] Bijone, M. (2016). A Survey on Secure Network: Intrusion Detection & Prevention Approaches. *American Journal of Information Systems*, 4(3), 69-88.
 - [52] Ashfaq, R. A. R., Wang, X. Z., Huang, J. Z., Abbas, H., & He, Y. L. (2017). Fuzziness based semi-supervised learning approach for intrusion detection system. *Information Sciences*, 378, 484-497
 - [53] Htun, P. T., & Khaing, K. T. (2012). Anomaly Intrusion Detection System using Random Forests and k-Nearest Neighbor. *Probe*, 41102(4107), 2377.
 - [54] <http://www.cs.waikato.ac.nz/ml/weka/>
 - [55] Weiss, S. M., & Indurkha, N. (1998). *Predictive data mining: a practical guide*. Morgan Kaufmann.
 - [56] Ahmim, A., & Ghoualmi-Zine, N. (2013). A new fast and high performance intrusion detection system. *International Journal of Security and Its Applications*, 7(5), 67-80.
 - [57] Kim, J., Shin, N., Jo, S. Y., & Kim, S. H. (2017). Method of intrusion detection using deep neural network. In *Big Data and Smart Computing (BigComp), 2017 IEEE International Conference on* (pp. 313-316). IEEE.

AUTHOR PROFILE

Ismail Keshta is an assistant professor in the Department of Computer and Information Technology at Dammam Community College (DCC), King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia. He received the B.Sc. and M.Sc. degrees in computer engineering and the Ph.D. degree in computer science and engineering from the King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia, in 2009, 2011, and 2016, respectively. He was a Lecturer with the Computer Engineering Department, KFUPM, from 2012 to 2016. Prior to that, in 2011, he was a Lecturer with Princess Nora Bint Abdulrahman University (PNU) and Imam Muhammad ibn Saud Islamic University (IMAMU), Riyadh, Saudi Arabia. His research interests include software process improvement, modeling, and intelligent systems.

An Enhanced Approach of Clustering Protocol to Minimize Energy Holes in Wireless sensor networks(WSN)

Chinmaya Kumar Nayak^{#1}, Satyabrata Das^{*2}

[#] Department of CSE & IT, Veer Surendra Sai University and Technology(VSSUT)
Burla, Odisha, India

¹ cknayak85@gmail.com

^{*} Department of CSE & IT, Veer Surendra Sai University and Technology(VSSUT)
Burla, Odisha, India

² teacher.satya@gmail.com

Abstract—The Energy hole problem is a major problem of data collection in wireless sensor networks. The sensors near the static sink serve as relays for remote sensors, which reduce their energy rapidly, causing energy holes in the sensor field. This project has proposed a customizable mobile sink based adaptive protected energy efficient clustering protocol (MSAPEEP) for improvement of the problem of energy holes along with that we also characterize and made comparison with the previous existing protocols. A MSAPEEP uses the adaptive protected method (APM) to discover the best possible number of cluster heads (CHs) to get better life span and constancy time of the network. The effectiveness of MSAPEEP is compared with previous protocols; specifically, low energy adaptive clustering hierarchy (LEACH) and mobile sink enhanced energy efficient PEGASIS based routing protocol using network simulator(NS2). Examples of simulation result show that MSAPEEP is more reliable and removes the potential of energy hole and enhances the stability and life span of the wireless sensor network(WSN).

Keywords: WSN, protected procedure, clustering protocols, mobile sink, energy hole problem.

I. INTRODUCTION

The Wireless Sensor Network (WSN) usually consists of a large number of costs in the surrounding environment, such as heat, pressure, vibration, appearance of objects, and so on. The concept of wireless sensor networks is based on a simple comparison:

Sensing + Processing + Communication = Thousands of potential applications

The measured capacity and procedures are then forward to a stationary network sensor. So, many clustering protocols have been particularly planned for WSNs to improve aggregation mechanism. These protocols differ significantly on the sharing system of nodes, network and radio model and network design. The difficulty with these protocols is the use of static dissipaters. Streaming directly into the sink does not guarantee a balanced load sharing of power between the distributions of energy load between sensors in the WSN and

thus increase network life span. The efficiency of WSNs is based on their sensory eminence, flexibility, coverage, etc., which they can provide. WSNs of course becomes the first choice when it comes to remote and dangerous deployment. The ultimate purpose of such WSNs dispersed in the aforementioned critical environments is often to provide survey data from the node sensors to the nozzle sink and then perform further analysis at the dive node. Data collection becomes an important factor in determining the performance of these WSNs.

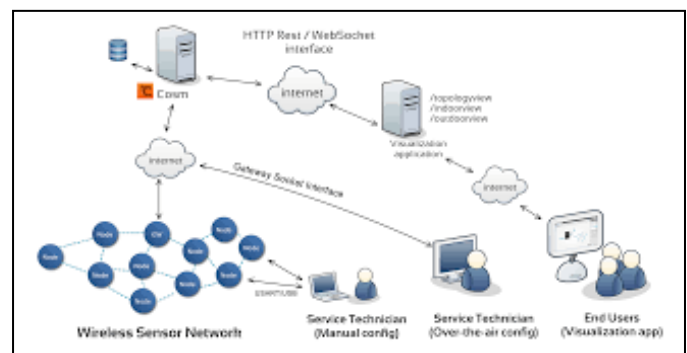


Fig. 1 An example of a WSN

II. CORRELATED WORK

An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

Many of the typical WSN clustering protocols, which consist of static nodes of sensors and a static dissipater, appeared in the literature. The Energy Efficiency Adaptive Accumulation Group (LEACH) is the first collection protocol. In LEACH, CH collects data from the sensors in its group and passes the data directly to the sink. The LEACH protocol problem is the random selection of CH. LEACH requires the user to specify the desired CH chance that he uses to decide if a node becomes a CH or less. The Mobile Synchronous Base (MSRP) protocol was deal with

to extend network life to WSNs grouped. In MSRP, the sink moves to the CH that has more energy in the heap network to collect the data they have collected. A new LEACH optimization clustering algorithm has been introduced with sinks and node assignments. This algorithm combines the use of the LEACH algorithm, mobile wash basin and appointment points to maintain the benefits of the LEACH algorithm and to improve the CH selection process.

It also reduces energy consumption in WSN compared to traditional LEACH, especially when the network is large. Mobile Roundup The advanced PEGASIS (MIEEPB) protocol for energy efficiency has been introduced. MIEEPB introduces the sink to the multi-chain model and divides the sensor field into four regions, thus reaching small chains and reducing load on the main joints. The sink moves along its trajectory and remains for some time in place in each region to ensure data collection. The removable sink in existing routing protocols always follows a certain trajectory and stops in fixed positions

This makes the sensors close to their fixed attitude to eradicate their energy faster than other joints. So in this document, we use a cable-controlled sink to minimize the distributed energy of all sensor nodes. In this case, the sensors surrounding the sink change over time, allowing all network sensors to act as a relay of data in the mobile sink and thus balancing the load between all the joints.

III. MOBILE SINK

The hole in the hole in the hole leads to an early grate disconnection and then the sink is isolated from the rest of the grid due to the death of the neighbors, while most of the sensor joints are still alive and fully functional. The use of the sink's mobility has been widely accepted as an effective way to mitigate the power cut off in WSN and to extend the life of the network by avoiding excessive overload on nodes that are near the sink. Pooling algorithms can actually arrange network joints and use a controlled sink to solve the power hole problem. However, finding the optimal number of CH and the optimal mobile trajectory for the mobile sink are non-deterministic polynomial problems - hard times (NP-hard). The sink is an important component of a WSN as it acts as a gateway between the sensor nodes and the end user.

The mobile sink has many advantages, such as increasing the security of WSNs. Since the position of mobile wash basins varies over time, harmful users are hard to know about its location and damage it. Therefore,

removable handheld can be useful for safe applications such as medical assistance, discovery of targets, and disclosure of battlefield interventions. In addition, the mobile sink improves network life and the rate of decline of the package. When a static sink is far from the sensor range or the sensor field is so large that most joints require a lot of steps to reach the sink, a considerable amount of rewind power is consumed during transmission so as to accelerate significant exhaustion of joints. However, mobile heat sink stores energy while data is transmitted in fewer steps. Thus, the number of packets down is reduced due to the sink movement closer to the sensor joints in the sensor field. In addition, the mobile dispenser improves network connectivity and eliminates power holes by balancing data routing between sensors.

IV. CLUSTERING

Naturally, grouping sensor nodes into clusters has been widely adopted by the research community to satisfy the above scalability objective and generally achieve high energy efficiency and prolong network lifetime in large-scale WSN environments. The corresponding hierarchical routing and data gathering protocols imply cluster-based organization of the sensor nodes in order that data fusion and aggregation are possible, thus leading to significant energy savings. In the hierarchical network structure each cluster has a leader, which is also called the cluster head (CH) and usually performs the special tasks referred above (fusion and aggregation), and several common sensor nodes (SN) as members. The cluster formation process leads to a two-level hierarchy where the CH nodes form the higher level and the cluster-member nodes form the lower level. The sensor nodes periodically transmit their data to the corresponding CH nodes. The CH nodes aggregate the data (thus decreasing the total number of relayed packets) and transmit them to the base station (BS) either directly or through the intermediate communication with other CH nodes. However, because the CH nodes send all the time data to higher distances than the common (member) nodes, they naturally spend energy at higher rates. A common solution in order balance the energy consumption among all the network nodes, is to periodically re-elect new CHs (thus rotating the CH role among all the nodes over time) in each cluster. The BS is the data processing point for the data received from the sensor nodes, and where the data is accessed by the end user. It is generally considered fixed and at a far distance from the sensor nodes. The CH nodes actually act as gateways between the sensor nodes and the

BS. The function of each CH, as already mentioned, is to perform common functions for all the nodes in the cluster, like aggregating the data before sending it to the BS.

In some way, the CH is the sink for the cluster nodes, and the BS is the sink for the CHs. Moreover, this structure formed between the sensor nodes, the sink (CH), and the BS can be replicated as many times as it is needed, creating (if desired) multiple layers of the hierarchical WSN (multi-level cluster hierarchy). In clustering, the sensor nodes are partitioned into different clusters. Each cluster is managed by a node referred as cluster head (CH) and other nodes are referred as cluster nodes. Cluster nodes do not communicate directly with the sink node. They have to pass the collected data to the cluster head. Cluster head will aggregate the data, received from cluster nodes and transmits it to the base station. Thus minimizes the energy consumption and number of messages communicated to base station. Ultimate result of clustering the sensor nodes is prolonged network lifetime network. It is the bridge (via communication link) between the sensor network and the end user. Normally this node is considered as a node with no power constraints. Cluster: It is the organizational unit of the network, created to simplify the communication in the sensor network. There are many types in clustering techniques used in wireless sensor network. After these techniques wireless sensor networks emerged as a best network for communication field.

A. Asymmetrical cluster used for WSN

The approximate sensor nodes in the base station consume more power because the network traffic is closer to the base station [14]. Therefore, the nodes die closer to the base station in time. To balance the energy consumption throughout the network, unequal classification methods are introduced. To maintain more power for intermittent data transmission, the network is divided into groups of unequal size and the groups closest to the base station are smaller than the base station. This article presents a complete study of the non-linear clustering algorithm for wireless sensor networks. The summarization and categorization algorithms are based on the group title and the selection of the duration of the network. The most commonly used non-uniform grouping algorithm is selected for comparison according to different properties

1) Energy-competent asymmetrical cluster (ECAC)

[6] is a competitively distributed decentralized grouping algorithm where group headings are chosen based on their neighbor's high residual energy and their distance to base station (BS). To solve the corrosion problem, the ECAC cluster buttons

spread to an uneven size and the closest grouping to the base station is smaller than those moving away from the BS because the node could not communicate directly with the BS at a limited distance. Each node has proved competitive. This competitive area reduces the distance to the base station. Consequently, clusters closest to the base station are smaller, so CH consumes less energy during communication within the group and can save more energy for group communication. The ECAC algorithm is also a probabilistic grouping algorithm because each node in each group creation cycle delivers a random number between 0 and 1 when it is decided to participate in the selection of groups. If the sensor node decides to join the cluster head selection, it becomes a preliminary group header. The initial group headline of local areas is part of the actual group headline. Competition is based on the remaining energy of each preliminary group headline. After the group header is selected, the remaining sensor nodes are connected to the nearest grouping.

2) Multihop direction-finding procedure with asymmetrical cluster (MDPAC)

[11] Choose CH cycles with high residual energy. Three steps, such as cluster installation, multi-environment routing between servers, and data transfer for each round. The data transfer step takes more time than the other two phases to reduce the total number of nodes. Initially, each node in the sensor network is responsible for collecting information from neighbors by sending HELLO messages to neighbors. First of all, all nodes are in an unknown state. The node is selected as CH if it has the most energy among all neighboring nodes and sends HEAD MSG to the nearest node in order to construct the group. Depending on the signal received, each neighbor calculates the distance $d(i, j)$ to BS based on the competing beam. If the node is not in the range of any CH and all its neighbors with a higher residual energy than it joins in other groups, the node is passively selected as CH. To mitigate the hot air defect, MDPAC accepts multi-channel data transmission and builds a blocking tree rooted in group B to save energy. A node with a minimum cost is treated as a parent node between all adjacent areas. Data transfer begins after the inter-server tree has been erected, and each node transmits data from the sensor to the CH at the specified transfer time. CH collects data packets in one packet and sends data to the master code that sends the packet received to the BS. The next round begins after a certain time. MDPAC outperforms a similar group version; extend the network's lifetime by 34.4%.

3) Asymmetrical Hierarchical Energy Competent disseminated cluster (AHECD)

[5] is an uneven grouping algorithm, extended with HECD. Uneven size is created based on the CH distance from the BS station. The radius of the ECAC competition formula [6] is used to create a smaller group closer to the BS. The amount of movement in the group decreases much closer to the BS, creating groups of unequal sizes. AHECD assumes the following assumptions about the node: (i) all nodes are homogeneous in terms of energy, communication and processing capabilities; (ii) each node is identified by a unique identifier; (iii) nodes may transmit at different power levels depending on the distance of the receivers; (iv) the nodes are not mobile and therefore remain stable after the evenly distributed implementation process; (v) communication nodes may determine the distance between them 1; (vi) all nodes know the distance from the base station. The BS station is received from the sensor network without energy

problems and is considered a nod to broad communication and computer skills. BS is not portable. The data captured in the group is strongly correlated; therefore it can be merged before sending to the base station. The problem of hot spots is effectively limited in AHECD as clusters of the same size and balances the energy consumption between sensor nodes in the network.

4) Energy Competent Disseminated Asymmetrical Cluster (ECDAC)

[12] is a decentralized, non-uniform grouping algorithm, where the group head can be divided by waiting times. The waiting time is measured by the residual energy parameters, the neighboring node number. Each button sends an advertisement message to calculate the number of neighboring nodes (NN) of a 1-hop series to calculate the distance from the base station. The base station gives each node a value that decides on time to create groups. ECDAC takes into account the node code area of the node and covers the entire network. The wait time of each sensor node is synchronized with the time of the node. If it reaches 0, the node is defined as CH. CH sends a HELLO message to neighbor nodes. The frequency of the HELLO message conversation depends on the distance to the BS station, the number of neighbor nodes, and the remaining CH energy. The neighbor node stores the CH shape on the weight table and changes the status as a member of the group. The member node sends a response message from their information to CH. During configuration, the next time the network node's standby time is determined. Compared to the EUE, energy consumption in ECDAC improved to 24.2%.

5) Energy-motivated Asymmetrical cluster (EMAC)

[10] Uneven competition areas are used in EMAC nodes to build uneven groups. Clusters from a further distance from the BS station are smaller to maintain energy for data transmission over long distances. Therefore, the energy consumption of the cluster head is effectively balanced. The cluster head can be rotated at the energy level of the cluster head to reduce unnecessary energy consumption. Each node acts as one cluster head throughout the lifetime of the network. In this way, EMAC reduces additional costs and ensures high energy efficiency. In this article, the energy level is calculated exactly when you rotate the cluster head, based on the assumption that the cluster head is a BS hopping connection. However, individual assumptions cannot be appropriate for the actual situation. In the random competition system used to select a group, it is not easy to estimate the number of packets sent by the cluster head when calculating the energy threshold. Therefore, the proposed energy-based connection plan is not suitable for multifunctional networks, because the energy intensity is defined as very accurate.

6) Asymmetrical Cluster-base Direction-Finding (ACDF)

[8] To alleviate the problem of heating points, the nodes are grouped into unequal groups. It is intended for the transfer movement between groups, which consists of two parts, one of which is EMAC, to alleviate the problem of hot spots, and the other is a routing protocol for transfer between groups [6].

In ACDF, power consumption is maintained uniformly in all CHs by reducing the number of nodes in the group with high load relays near the base station. Initially, the initial CH is chosen at random to compete for the final CH. Each initial group heading has a series of competitions. Different ranges of competence are used to produce a group of unequal sizes. Finally, one CH is allowed in each competition series. After selecting CH, each broadcast message on the network. Each node selects its closest CH with the highest reception delay and sends a message about the cluster connection to the nearest group header. The appearance of the sensor node is then built. ACDF assumes that the approximate distance from one sensor to another depends on the strength of the reception signal. In a real environment, the error occurs as a result of the noise.

7) Asymmetrical Cluster Dimension (ACD)

[7] An asymmetrical grouping model based on the Asymmetrical Group Dimension (ACD) to balance the energy consumption of group leaders due to the intense group diversion. groups In LUW, nodes N are randomly distributed in a circular area with radius R. BS is located in the center of the observed area and receives all information collected in CH. Data transfer can be done using multihops. Each CH selects the nearest CH channel to send shared data to the base station. In general, CH creates similar to the unbalanced base station power consumption of all CHs. Maintain more homogeneous energy consumption within CH by turning group heads in each group. In the ACD, the CH channel positions are predetermined to arrange the CHs in a symmetric concentric circle relative to the base station. Each cluster node is collected in the Voronoi region around CH. It provides a layered network in which each layer contains a number of clusters. ACD assumes that all conglomerates in the layer have the same size and shape, but differ from each other. In multidecked networks, the ACD was 10-30% better than the existing equal grouping models

| cluster Proce dures | parameter use to choose CH | | | | | Connect iv ity of CH to BS | Nod e typ e | Mode of CH Selecti on | Numb er of Nodes in cluster | Ener gy Efficie ncy | Balanc ed Cluste r |
|---------------------------|----------------------------------|----|--------|----|--------|-------------------------------------|----------------------|--------------------------------|---|------------------------------|-----------------------------|
| | R E | BS | N N | SH | M H | | | | | | |
| MDPA C | ✓ | ✗ | ✓ | ✗ | ✓ | | Homo | Neighbor | Unequal | Moderat e | No |
| ECAC | ✓ | ✗ | ✗ | ✗ | ✓ | | Homo | Probabilis tic | Unequal | Good | Yes |
| ECDA C | ✓ | ✗ | ✓ | ✓ | ✓ | | Hetero | Probabilis tic | Unequal | Moderat e | No |
| ACD | ✓ | ✗ | ✗ | ✗ | ✓ | | Homo | Probabilis tic | Equal | Good | Yes |
| ACDF | ✓ | ✗ | ✗ | ✗ | ✓ | | Homo | Probabilis tic | Equal | Good | Yes |

| | | | | | | | | | | |
|-----------|---|---|---|---|---|--------|------------|---------|---------|----|
| EMAC | ✓ | ✗ | ✗ | ✗ | ✓ | Hetero | Probabilis | Unequal | Moderat | No |
| AHEC D | ✓ | ✗ | ✗ | ✗ | ✓ | Homo | Probabilis | Unequal | Moderat | No |

V. PROPOSED SYSTEM

We propose a customizable mobile sink based adaptive protected energy efficient clustering protocol (MSAPEEP) to alleviate the energy holes. MSAPEEP uses the adaptive protected method (APM) to find the locations of the mobile sink and the optimum number of CHs and their break locations based on minimizing the total dissipated energy in communication process and overhead control packets of all sensor nodes within the network. In our protocol, we use a controlled mobile sink that guided based on minimizing the dissipated energy of all sensor nodes. The sensor field is divided into R equal size regions to conserve energy since data is transmitted over fewer hops. This reduces the number of dropped packets and delay that packet needs to reach to the sink because the mobile sink moves along the break path and stops at the break location closer to the sensor nodes in each region in the sensor field.

A. Prepare Phase

In this phase, the sink initializes the network by defining the number of nodes, the data packet size, the control packet size, the size of sensor field and the parameters of the radio model. Then the sink divides the sensor field into R equal size regions; where N/R nodes are deployed randomly in each region. After that, the sink initially moves to center of each region and requests the ID, position and Eo of all sensors in each region. The connectivity between nodes and the sink is always satisfied, because the communication radius for each node is assumed to be larger than the coverage radius.

B. Set-Up Phase

After initialization, the mobile sink goes to center of r th region ($r = 1, 2, \dots, R$) and uses APM to find its break location and locations of the optimum CHs based on the minimization of the total dissipated energy in communication. Then the mobile sink assigns the members nodes of each CH. If a sensor is close to the sink than any CH in this region, this node will communicate directly to the sink. Once CHs are selected and members of each CH are assigned, the sink broadcasts two short messages. The first one is sent to the selected CHs to inform each one by IDs of its members. While the second message that contains CH's ID and logic 0 is sent to member nodes to inform each one where will join. Based on the received messages from the sink, each CH in r th region creates the TDMA schedule by assigning slots to its member nodes and informs these nodes by the schedule. The TDMA schedule is used to avoid intra-cluster collisions and reduce energy consumption between data messages in the cluster

and enable each member of the radio equipment off when not in use.

C. Steady State Phase

After finding the locations of the CHs and the sojourn location of the mobile sink in a region r , the sink moves to its sojourn location and wakes up the sensor nodes in this region, while the rest nodes in other ($R-1$) regions are sleep. The nodes start sensing the data; then each sensor sends its data to its CHs or the sink if it is close to the sink than CH according to the TDMA schedule. Each cluster communicates using different CDMA codes in order to reduce interference from nodes belonging to other clusters. Once each CH received the sensed data from its member nodes, it performs signal processing functions to aggregate the data into a single packet. Then, CHs send their packets to the sink. After certain time called sojourn time, the sink moves at a certain speed along the mobility path to the next region ($r+1$) to perform clustering and collects data from the sensors in this region. This process is repeated until the sink visits all R regions in the sensor field to guarantee complete data collection. When the sink finishes its round, it again goes back to first region to begin a new round.

Advantages

- ☐ Reduce the dropped packets
- ☐ Decrease the time delay
- ☐ Provide the efficient packet delivery
- ☐ Reduce energy consumption

VI. RESULTS AND ANALYSIS

The proposed protocol extends the stability period and improves the network lifetime as compared to MIEEPB protocol and rendezvous protocol for the three mobility path patterns respectively. This means that the proposed protocol is more energy-efficient than the other protocols, because it allows for nodes to work with full functionality for long time due to the higher residual energy of the sensor nodes in the network. Furthermore, the residual energy of all nodes in the network for the proposed protocol decreases more slowly than other protocols when the number of rounds of the nodes increases. Using the mobile sink with APM (Adaptive Protected Method) eliminates the energy holes and out performs the other protocols in terms of stability period, throughput, packet delivery ratio and in the lifetime.

TABLE I. SIMULATION SETUP PARAMETERS

| S.No. | parameter | Ideals |
|-------|-------------------|-----------------|
| 1. | Region | 1000 m × 1000 m |
| 2. | Quantity of Nodes | 35, 50, 100 |
| 3. | Node Mobility | No |

| | | |
|----|--|---|
| 4. | Traffic | CBR (bits/sec) |
| 5. | Initial energy of nodes | 2 Joules |
| 6. | Sink node position | (75, 175) |
| 7. | ϵ_{fs} (free space model energy consumption) | 10 pJ/bit/m ² |
| 8. | ϵ_{mp} (multi path model energy consumption) | 0.0013 pJ/bit/m ⁴ |
| 9. | Cross over point do | $(\epsilon_{fs}/\epsilon_{mp})^{1/2}$ m |

Performance Metrics

The following metrics are used to evaluate the performance of the proposed protocol

- **Number of Alive Nodes per Round:** The number of nodes that have not yet expended all of their energies.
- **Network Lifetime:** The time interval from the start of network operation until the death of the last alive sensor.

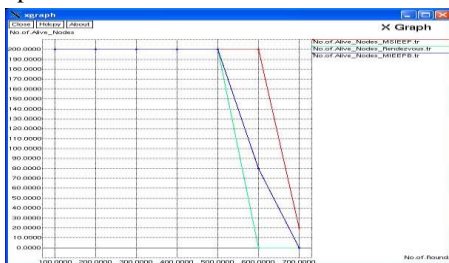


Fig 2. Network Lifetime

- **Stability Period:** The time interval from the start of network operation until the death of the first sensor.
- **Throughput:** It measures the total rate of data sent over the network, including the rate of data sent from CHs to the sink and the rate of data sent from the nodes to their CHs.
- **Packet delivery ratio:** It measures the ability of a protocol to deliver packets to the destination. It is the ratio of the number of packets that are successfully delivered to the destination to the total number of packets that are sent.

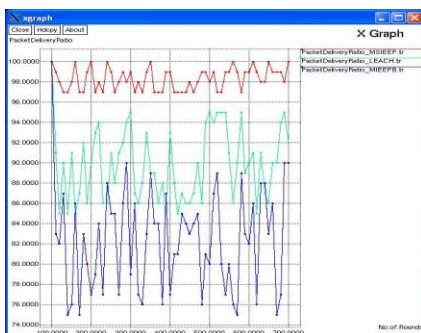


Fig 3. Packet Delivery Ratio

- **Packet drop ratio:** It measures the robustness of protocol and is calculated by dividing the total number of dropped packets by the total number of transmitted packets.

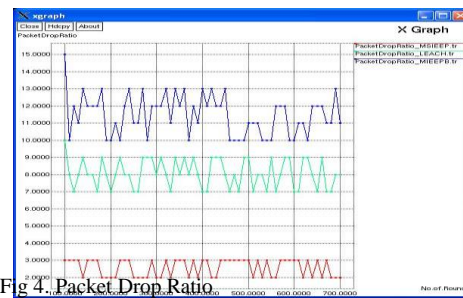


Fig 4. Packet Drop Ratio

- **Packet delay:** The time required by a packet to reach from source to destination. It is calculated by dividing the distance from source to destination by the speed of light

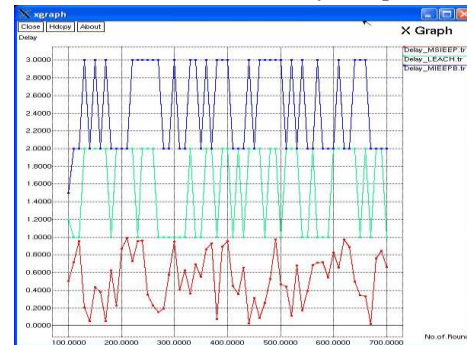


Fig 5. Packet Delay

SIMULATION RESULTS

TABLE 2. SIMULATION RESULTS

| Parameters | LEACH and rendezvous Protocol | MSIEEP Protocol | Percentage of Improvement (%) |
|---------------------------|-------------------------------|-----------------|-------------------------------|
| Packet drop ratio(%) | 11.42 | 0.57 | 95 |
| Packet delay(μ sec) | 11.76 | 8.45 | 28 |
| Packet delivery ratio (%) | 75.06 | 98.06 | 23 |

From these results, it is noticed that the packet drop ratio and the packet delay increase as the number of nodes increases for all protocols. Sending more packets for high node degree

cases increases the chances of more dropped packets due to increased congestion at the receiver end which causes buffer overflow and thus leading to dropped packets and higher packet delay. Introducing mobility to sink and dividing the sensor field into small size regions in the proposed protocol improve the probability of packet drop and packet delay compared to the other protocols. Moreover, this increases the robustness and the ability of the proposed protocol to deliver packets to the destination.

VII. CONCLUSIONS

A customizable mobile sink based adaptive protected energy efficient clustering protocol (MSAPEEP) has been introduced to eliminate the Energy hole problem and further enhance the life span and the stability period of WSNs. In addition, this protocol uses the Adaptive Protected Method (APM) and the optimum number of cluster heads and their positions based on the minimization of distributed energy in the communication control packets and the control of all sensor nodes in the sensor field. The simulation results showed that the proposed protocol is more reliable and energy-efficient than other existing protocols; i.e. the LEACH, LEACH-GA, A-LEACH, Rendezvous and MIEEPB protocols. It also exceeds previous protocols in terms of longevity, stability period, packet release ratio, and packet delay. Future work can be done using the road planning algorithm for the mobile washbasin. In which the delay will be reduced and will further improve the distribution ratio of the wireless sensor network packet

REFERENCES

- [1] Abbasi.A.A and Younis.M, (Oct 2007)—A survey on clustering algorithms for wireless sensor networks, *Comput. Commun.*, vol. 30, nos. 14–15, pp. 2826–2841.
- [2] Zhou.Q, Cao.X, Chen.S, and Lin.G, (Dec 2009) —A solution to error and loss in wireless network transfer, *in Proc. Int. Conf. WNIS, Shanghai, China*, pp. 312–315.
- [3] Singh.S.K, Singh.M.P, (Nov 2010) and D. K. Singh, —Routing protocols in wireless sensor networks—A survey, *Int. J. Comput. Sci. Eng. Surv.*, vol. 1, no. 2, PP.63–83.
- [4] Liu.J.L and Ravishankar.C.V, (Apr 2011) —LEACH-GA: Genetic algorithm-based energy-efficient adaptive clustering protocol for wireless sensor networks, *Int. J. Mach. Learn. Comput.*, vol. 1, no. 1, 79–85.
- [5] Lu.X, Ding.Y, and Hao.K, (Jan 2011) —Immune clonal selection algorithm for target coverage of wireless sensor networks, *Int. J. Model., Identificat., Control*, vol. 12, nos. 1–2, pp. 119–124.
- [6] Abo-Zahhad.M, Ahmed.S.M, Sabor.N, and Al-Ajlouni.A.F, (Feb 2012)—A new method for fastening the convergence of immune algorithms using an adaptive mutation approach, *J. Signal Inf. Process.*, vol. 3, no. 1, pp. 86–91.
- [7] Shinghal.K, Noor.A, Srivastava.N, and Singh.R, (2011)—Power measurements of wireless sensor network node, *Int. J. Comput. Eng. Sci.*, vol. 1, no. 1, pp. 8–13.
- [8] Gupta.S.K, Jain.N, and Sinha.P, (Aug 2013) —Energy efficient clustering protocol for minimizing cluster size and inter cluster communication in heterogeneous wireless sensor network, *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 2, no. 8, pp. 3295–3304, Aug. 2013.
- [9] Ahmad.A, Javaid.N, Khan.Z.A, Qasim.U, and Alghamdi.T.A, (Oct 2014) —(ACH)2: Routing scheme to maximize lifetime and throughput of WSNs, *IEEE Sensors J.*, vol. 14, no. 10, pp. 3516–3532.
- [10] Sandhu.M.M, Akbar.M, Behzad.M, Javaid.N, Khan.Z.A, and Qasim.U, (Nov 2014)—Mobility model for WBANs, *in Proc. 9th Int. Conf. BWCCA, Guangzhou, China*, Nov. 2014, pp. 155–160.
- [11] Mottaghi.S and Zahabi.M.R, (Feb 2014)—Optimizing LEACH clustering algorithm with mobile sink and rendezvous nodes, *AEU-Int. J. Electron. Commun.*, vol. 69, no. 2, pp. 507–514.
- [12] Chuang.P.J. and Jiang.Y.J., (Jun 2014) —Effective neural network-based node localisation scheme for wireless sensor networks. *IET Wireless Sensor Syst.*, vol. 4, no. 2, pp. 97–103.
- [13] Lin Zhao Zhibo Chen Guodong Sun School of Information Science and Technology, Beijing Forestry University, Beijing 100083, China —Dynamic Cluster-based Routing for Wireless Sensor Networks, *Journal of International*, Vol. 11, November 2014
- [14] Daisuke Takaishi, Hiroki Nishiyama, Nei Kato, and Ryu Miura, Graduate School of Information and Communications Technology, Tokyo Japan —Towards Energy Efficient Big Data Gathering in Densely Distributed Sensor Networks, *IEEE Transaction Volume 2*, No. 3, September 2014
- [15] F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, —A survey on sensor networks, *IEEE Commun. Mag.*, vol. 40, pp. 102–114, August 2002.
- [16] D. Puccinelli, M. Haenggi, —Wireless sensor networks: applications and challenges of ubiquitous sensing, *IEEE Circuits Syst. Mag.*, vol. 5, pp. 19–31, Sep 2005.
- [17] S. Olariu and I. Stojmenovic, —Design Guidelines for Maximizing Lifetime and Avoiding Energy Holes in Sensor Networks with Uniform Distribution and Uniform Reporting, *in Proc. IEEE INFOCOM' April 2006*, pp. 1–12.
- [18] N. Ahmed, S.S. Kanhere, and S. Jha, —The holes problem in wireless sensor networks: a survey, *in Proc. SIGMOBILE Mobile Computing and Communications Review 2005*, 2009, pp. 4–18.
- [19] N. Jabeur, N. Sahli, I.M. Khan, —Survey on Sensor Holes: A Cause-Effect-Solution Perspective, *in Elsevier Proc. 8th International Symposium on Intelligent Systems Techniques for Ad hoc and Wireless Sensor Networks (IST-AWSN) 2013*, pp. 1074 – 1080.
- [20] Khan, H. Mokhtar, and M. Merabti, —An overview of holes in wireless sensor networks, *in Proc. of the 11th Annual Postgraduate Symposium on the Convergence of Telecommunications, Networking and Broadcasting*, June 2010.
- [21] Q. Fang, J. Gao, and L.J. Guibas. —Locating and bypassing routing holes in sensor networks, *in Proc. INFOCOM 2004. 23th Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 4, March 2006, pp. 2458–2468.
- [22] S.-M. Jung, Y.-J. Han, and T.-M. Chung, —The concentric clustering scheme for efficient energy consumption in the PEGASIS, *in Proc. 9th Int. Conf. Adv. Commun. Technol.*, vol. 1, Phoenix Park, Korea, Feb. 2007, pp. 260–265.
- [23] S. Deng, J. Li, and L. Shen, —Mobility-based clustering protocol for wireless sensor networks with mobile nodes, *IET Wireless Sensor Syst.*, vol. 1, no. 1, pp. 39–47, 2011.
- [24] Jitendra Rajpurohit, Tarun Kumar Sharma, Ajith Abraham, Vaishali, Glossary of Metaheuristic Algorithms, *International Journal of Computer Information Systems and Industrial Management Applications*, 9(2017), pp 181–205.
- [25] Tarun Kumar Sharma, Millie Pant, (2017), Opposition-Based Learning Embedded Shuffled Frog-Leaping Algorithm, *Proceedings of First International Conference on Soft Computing: Theories and Applications*, pp 853–861.
- [26] Tarun K. Sharma, (2017), Performance Optimization of the Paper Mill using Opposition based Shuffled frog-leaping algorithm, *International Journal of Computer Information Systems and Industrial Management Applications*, 9(2017), pp 173–180
- [27] TK Sharma, M Pant, (2017), Distribution in the placement of food in artificial bee colony based on changing factor, *International Journal of System Assurance Engineering and Management*, 8(1), pp 159–172
- [28] TK Sharma, Pant Millie, (2017), Shuffled artificial bee colony algorithm, *Soft Computing*, 21(20), pp 6085–6104

Enhanced Hybrid Blowfish and ECC Encryption to Secure Cloud Data Access and Storage Policies

Jobandeep Kaur¹, Dr Vishal Bharti², Mr. Shamandeep Singh³

M. Tech (Scholar), Head of Department, Assistant Professor

Department of Computer Science and Engineering, Chandigarh, Gharuan

Punjab (INDIA)

Jobandeeprandhawa4@gmail.com¹, hod.ibmcse@cumail.in², shamandeep.cse@cumail.in³

Abstract – Encryption is the most important concept to enhance the security in cloud access policies. Encryption data in the cloud is the procedure of transforming or encrypted data or information before it's moved to cloud storage. Normally, in cloud service sources give encrypted services ranging from an encoding connection to limited encode sensitive information and provide encode key to decode the data as required. Several security problems and some of their solution are examined and are concentrating primarily in public security problems and their solutions. In this paper, we've implemented a hybrid approach, where access policies won't leak any privacy data and to enhance the security and performance parameters like decryption time, encryption time and accuracy and compared with existing performance parameters. Security is the main limitation while storing data over cloud server. The introduced approach is implemented appropriately even if the tenant could access the information all that would appear is gabble., Hijacking of sessions while accessing data, insider threats, outsider malicious attacks, data loss, loss of control, and service disruption. Therefore enhancing the security for multimedia data storage in a cloud center is of paramount importance.

Keywords – Role based access control, Encryption, Decryption, ECC, and Blowfish.

I. INTRODUCTION

Role based access control (RBAC) is a technique for controlling access to PC or system assets in view of the parts of individual clients inside a venture. In this unique circumstance, get to is the capacity of an individual client to play out a particular assignment, for example, see, make, or adjust a document. The idea of RBAC started with multi-client and multi-application on-line frameworks spearheaded in the 1970s. Clients can be effectively reassigned from one part to another. Parts can be allowed for new authorizations as new applications and frameworks are joined, and authorizations can be disavowed from parts as required [1]. Three basic principles of RBAC are:

- An individual must be allotted a specific part with a specific end goal to lead a specific activity, called an exchange.
 - A client needs a part approval to be permitted to hold that part.
 - Exchange approval enables the client to play out specific exchanges. The exchange must be permitted to happen through the part enrolment.
- Attribute based access control (ABAC) is model which develops from RBAC to think about extra ascribes notwithstanding parts and gatherings [2]. Managing and examining system get to is basic to data security. Access can and ought to be allowed on a need-to-know premise. With hundreds or thousands of workers, security is all the more effectively kept up by restricting pointless access to touchy data in view of every client's built up part inside the association [3]. Several benefits of RBAC are:
- *Reducing administrative work and IT support:* With RBAC, we can lessen the requirement for printed material and secret word changes when a representative is enlisted or changes their part. RBAC additionally serves to all the more effortlessly incorporate outsider clients into your system by giving them pre-characterized parts.
 - *Maximizing operational efficiency:* RBAC offers a streamlined approach that is coherent in definition. Every part can be lined up with the hierarchical structure of the business and clients can carry out their employments all the more effectively and self-governing.
 - *Improving Compliance:* All associations are liable to government, state and neighbourhood directions. This is critical for human services and money related foundations, which oversee bunches of touchy information, for example, PHI and PCI information.

Cloud encryption is an administration offered by distributed storage suppliers whereby information, or content, is changed utilizing encryption calculations and is then put on a capacity cloud. Cloud encryption

is the change of a cloud benefit client's information into figure content.

The cloud encryption capacities of the specialist organization need to coordinate the level of affectability of the information being facilitated [4]. Cloud computing depends on five traits:

- *Multi-tenancy (shared assets)*: Cloud processing depends on a plan of action in which assets are shared (i.e., various clients utilize a similar asset) at the system level, have level, and application level.
- *Massive scalability*: Cloud processing gives the capacity to scale to countless frameworks, and the capacity to hugely scale data transfer capacity and capacity space.
- *Elasticity*: Users can quickly increment and abatement their processing assets as required.
- *Pay as you utilized*: Users pay for just the assets they really utilize and for just the time they require them.
- *Self-provisioning of resources*: Users self-arrangement assets, for example, extra frameworks (handling capacity, programming, stockpiling) and system assets.

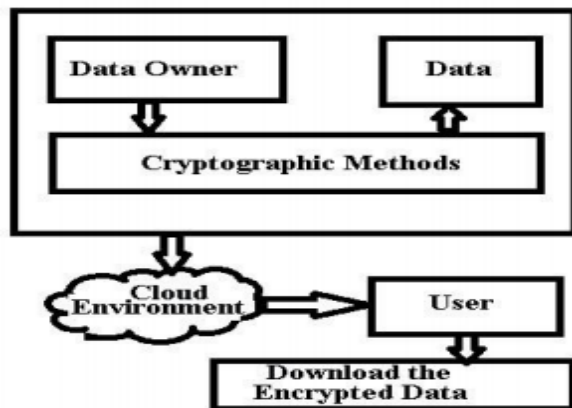


Figure 1 Cloud Strategy [4]

Encryption techniques can be applied to data on the drive or array, at the host or in the fabric. The fundamental segments of a cryptographic stockpiling administration which can be actualized by utilizing alternate systems, out of which, some were planned particularly for distributed storage. In the start of the Cloud Computing, normal encryption Technique like Public Key Encryption was connected. The progressed cryptographic strategies incorporates the underneath encryption techniques.

ECC is an open source encryption strategy in light of elliptic bend hypothesis that can be utilized to make speedier, littler, and more productive cryptographic keys. ECC creates keys through the properties of the elliptic bend condition rather than the conventional strategy for age as the result of vast prime numbers.

The innovation can be utilized as a part of conjunction with most open key encryption strategies, for example, RSA, and Diffie-Hellman. ECC was created by Certicom, a versatile e-business security supplier, and was as of late authorized by Hifn, a maker of incorporated hardware (IC) and system security items. Later several manufacturers have included help from ECC in their items [5].

Blowfish Algorithm is a symmetric square figure, laid out by Bruce Schneier in 1993, that can be reasonably used for encryption and safeguarding of data. Blowfish scrambles 64 bit obstructs with a variable length key of 128-448 bits. As per Schneier, Blowfish was outlined because of the followings goals: [6]

- Fast-Blowfish encryption rate on 32-bit microchips is 26 clock cycles for each byte.
- Compact-Blowfish can execute in under 5 kb memory.
- Simple-Blowfish utilizes just crude activity - s, for example, expansion, XOR and table look into, making its plan and execution basic.
- Secure-Blowfish has a variable key length up to most extreme of 448-piece long, making it both secure and adaptable.

In this paper, we've implemented a hybrid approach, to calculate data access policies and to enhance the performance of framework and calculate performance parameters like: encryption, decryption and accuracy.

In this section we've discussed the encryption techniques. We also reviewed the techniques used in our frame work to enhance the performance. In section II, we reviewed and analyzed the existing work done in this get better idea of field and present and future trends in cryptographic strategies. In section III, we have compared the feature of encryption techniques as well as the encryption algorithms. In Section IV, design and implementation of proposed methodology of framework is explained. Lastly, in section V, all results are explained.

II. RELATED WORK

Kan Yang, et al., (2017) [7] proposed a proficient and fine-grained huge data get to control access with protection safeguarding strategy. Step by step an instruction to control the entrance of the tremendous measure of huge information turns into an extremely difficult issue, particularly when enormous information are put away in the cloud. CP-ABE (Cipher text-Policy Attribute based Encryption) is promising encryption procedure that empowers end-clients to scramble their information under the entrance strategies characterized over a few traits of information shoppers and just permits information customers whose qualities fulfil the entrance

approaches to unscramble the information. In CP-ABE, the entrance approach is appended to the cipher text in plaintext shape, which may likewise release some private data about end-clients. Existing strategies just mostly shroud the property estimations in the entrance approaches, while the characteristic names are as yet unprotected. Particularly, they shroud the entire trait (as opposed to just its esteems) in the entrance arrangements. To help information unscrambling, we likewise outline a novel Attribute Bloom Filter to assess whether a trait is in the entrance approach and find the correct position in the entrance arrangement on the off chance that it is in the entrance strategy. Security examination and performance assessment explains that the strategy can preserve the protection from any LSSS get to arrangement without utilizing much overhead.

Qi Yuan, et al., (2015) [8] reviewed an issue of fine-grained information access control in distributed computing and proposed access control strategy to accomplish fine grainedness and execute the task of client denial effectively. The application programming in Cloud Computing and databases are moved to expansive incorporated server farms, where the administration of the information and administrations may not be completely reliable. This special worldview brings numerous new security challenges, which have not been all around fathomed. Information get to control is a compelling method to guarantee the huge information security in the cloud. The investigation comes about demonstrate that our plan guarantees the information security in distributed computing and decreases the cost of the information proprietor fundamentally.

Varsha S. Bandagar, et al., (2015) [9] outlined a cipher text-policy approach based encryption (ABE) plot to address an issue of absence of mechanisms to get control. In addition they proposed a safe, productive and fine grained information access control instrument for P2P cloud namely ACPC. In cloud computing, P2P storage by integrating method storage cloud is shaped to offer exceptionally accessible capacity administrations, bringing down the financial cost by abusing the storage room of partaking clients. Be that as it may, since cloud separates and clients are normally outside the put stock in space of information proprietors, distributed capacity cloud delivers new difficulties for information security and access control when information proprietors store touchy information for partaking in the put stock in area. Characteristic based encryption plot with effective client denial the execution assessment processing overhead diminishing when the look at the before client renouncement information proprietor and server.

Mohamed Nabeel, et al., (2012) [10] discussed the disadvantages of various methodologies on the basis of known cryptographic systems in tending issues and existing 2 methodologies that address those disadvantages with various trade-offs. With numerous practical advantages of distributed computing, numerous associations have been thinking about moving their data frameworks to the cloud. In any case, a critical issue openly mists is the means by which to specifically share information in view of fine-grained quality based access control approaches while in the meantime guaranteeing secrecy of the information and protecting the security of clients from the cloud.

Bilel Zaghdoudi, et al., (2016) [11] proposed an approach in view of DHT toward get to control for specially appointed MCC and Fog registering. They depend on Chord DHTs to make a versatile, nonexclusive and powerful access control arrangement. They utilize reproductions to assess the exhibitions of the proposition. They centred on an arrangement of measurements to gauge the overhead of the framework. They considered a variable system estimate, a variable dependable hubs rate and distinctive hash work as recreation parameter. They got comes about show satisfactory overhead for generally normal systems sizes. Re-enactments demonstrate that every one of the measurements increment with the hubs number and the quantity of dependable hubs.

Ying-Qian Zhang, et al., (2015) [12] proposed fresh image encryption calculation which depends on the spatiotemporal non-contiguous coupled guide cross sections. The arrangement of non-neighbouring coupled guide cross sections has more exceptional cryptography includes in elements than the strategic guide or coupled guide grids does. In the proposed picture encryption, they utilize somewhat level pixel stage methodology which empowers bit planes of pixels permute commonly with no additional storage room. Recreations have been done and the outcomes exhibit the unrivalled security and high effectiveness of the proposed calculation.

III. OVERVIEW OF ENCRYPTION TECHNIQUES

Encryption techniques can be applied to data on the drive or array, at the host or in the fabric. The fundamental segments of a cryptographic stockpiling administration which can be actualized by utilizing alternate systems, out of which, some were planned particularly for distributed storage. Several progressed encryption techniques are:

Table 1 Feature Comparison of Encryption Techniques [13]

| Technique | Fine Grained access control | Computation Overhead | User revocation efficiency | Scalability/ efficiency | Collision resistance | Attributes Association | Access Policy Association |
|-----------|-----------------------------|----------------------|----------------------------|-------------------------|----------------------|------------------------|---------------------------|
| IDE | Low | Avg | Avg | Avg | Low | With Cipher | With Key |
| ABE | Low | Avg | Avg | Avg | Low | With Cipher | With Key |
| KP-ABE | Avg | Mostly Overhead | Low | Avg | Above Average | With Cipher | With Key |
| CP-ABE | Avg | Avg | Low | Avg | Good | With Key | With Cipher |
| HIBE | Comparative Low | Mostly Overhead | - | Better | Good | - | - |
| HABE | High | Overhead | Avg | Above Avg | Good | With Key | With Cipher |
| MA-ABE | Better | Avg | High | High | Good | With Cipher | With Cipher |

Table 2 Comparison of various algorithms based on different parameters [14]

| PARAMETERS | DES | 3DES | AES | RSA | BLOWFISH |
|-------------------|---|--------------------|--|---|---|
| Development | In early 1970 by IBM and published in 1977 | IBM in 1978 | Vincent Rijmen, Joan Daeman in 2001 | Ron Rivest Shamir & Leonard Adleman in 1978 | Bruce Schneier in 1993 |
| Key length (Bits) | 64 (56 usable) | 168/112 | 128, 192, 256 | Key length depends on no. of bits in module | Variable key length i.e. 32 - 448 |
| Rounds | 16 | 48 | 10, 12, 14 | 1 | 16 |
| Block Size (Bits) | 64 | 64 | 18 | Variable block size | 64 |
| Attacks Found | Exclusive Key Search, Linear cryptanalysis, Differential analysis | Related Key attack | Key recovery attack, Side channel attack | Brute force attack, timing attack | No attack found to be successful against blowfish |
| Level of Security | Adequate Security | Adequate Security | Excellent Security | Good Security | Highly Secure |
| Encryption Speed | Very Slow | Very Slow | Faster | Average | Very Fast |

Table 3 Comparison of various algorithms based on different parameters.

| PARAMETERS | TWOFISH | THREEFISH | RC5 | ECC | IDEA |
|-------------------|------------------------|--|-------------------------------|---|------------------------------|
| Development | Bruce Schneier in 1998 | Bruce Schneier, Niels Ferguson, Stefan Lucks in 2008 | Ron Rivest in 1994 | Victor Miller from IBM and Neil Koblitz in 1985 | Xuejia Lai and James in 1991 |
| Key length (Bits) | 128, 192, 256 | 256, 512, 1024 | 0 – 2040 bits (128 suggested) | Smaller but effective key | 128 |
| Rounds | 16 | For 256 & 512 keys = 72 and for 1024 keys = 80 | 1 – 255 (64 suggested) | 1 | 8 |
| Block Size (Bits) | 128 | 256, 512, 1024 | 36, 64, 128 (64 suggested) | Stream size in Variable | 64 |
| Attacks Found | | | Co-relation attack | Doubling attack | Linear attack |
| Level of Security | Secure | Secure | Secure | Highly Secure | Secure |

IV. PROPOSED WORK

Security is the main limitation while storing data over cloud server. Various security threats in cloud computing are Data loss, Leakage of data, User's authentication, Malicious users handling, Wrong usage of Cloud computing and its services, Hijacking of sessions while accessing data, insider threats, outsider malicious attacks, data loss, loss of control, and service disruption. Therefore enhancing the security for multimedia data storage in a cloud centre is of paramount importance. Developing such an architecture which ensures the user that its data is secure is the main objective. Currently used approaches need some optimizations to increase the security and accuracy factor for storing and accessing the data among various users. Time consumption for decryption is also high. Management of various roles in the access policies is very time consuming to load and difficult to manage while working with large systems.

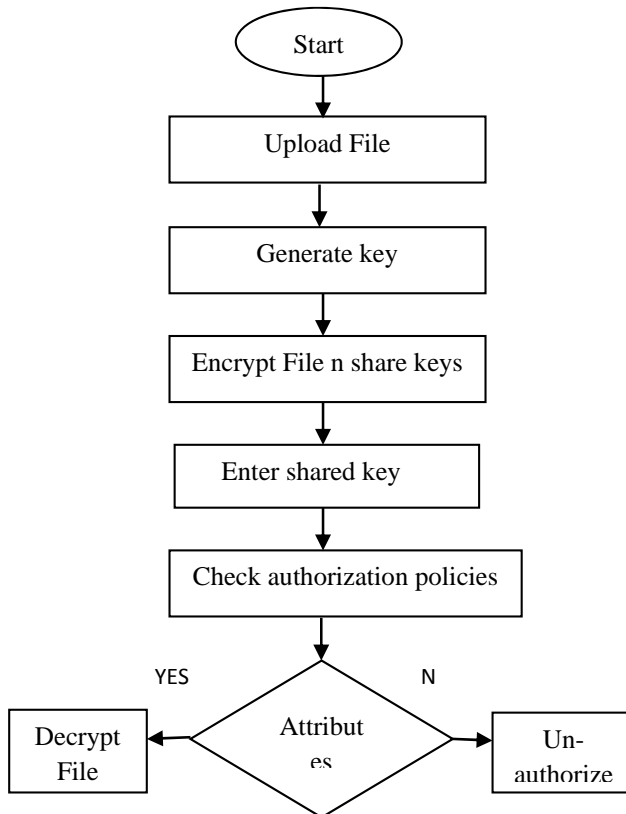


Figure 2 Proposed Flowchart

Security in access control mechanism is a challenging task. The overall process divided into various sub modules to find and optimize the working of storage and access policies. In this scenario the system get the file from use end and store in cloud repositories. The storage of data at on this platform is followed by encryption policies. The proposed architecture is used

to optimize the encryption scheme to eliminate the un-authorized access of user storage. The proposed architecture is a hybrid of two different algorithms asymmetric and symmetric. These two algorithms make the encryption scheme more secure and decrease the decryption probability. In this flow the user upload their files to store in cloud repositories and system extract all bytes from the uploaded data. Extracted bytes passed to the first step where the system generates keys for user authentications. After this process the keys passed to the encryption algorithm. In encryption module use the already generated keys to encrypt and make the process of data storage more secure. Various parameters are used to check the efficiency of the system and accuracy of the output files as compare to the original data.

Step1- Select master record from install message catch.

Step2- Select any photo from the nearby drive.

Step3- After choosing expert document select yield record to insert message.

Step4 - If the document ought to be packed at that point tap on check box pack.

Step5- If message ought to be scrambled at that point Click on checkbox encode message.

Step6- If the message ought to be concealed at that point compose message in message box and tap on go catch, at that point discourse will be show up with task is effective or not.

Step7- Close inserting message window by tapping on close catch.

Step8- To recovering scrambled, covered up, compacted message tap on recover message catch and select the yield document.

Step9- Tap on go catch and enter the scrambled secret word for recovering message.

V. RESULTS AND DISCUSSIONS

In this section, explained the encryption based results and comparison show in bar graph format. In proposed work, has implemented a hybrid approach to enhance the security.

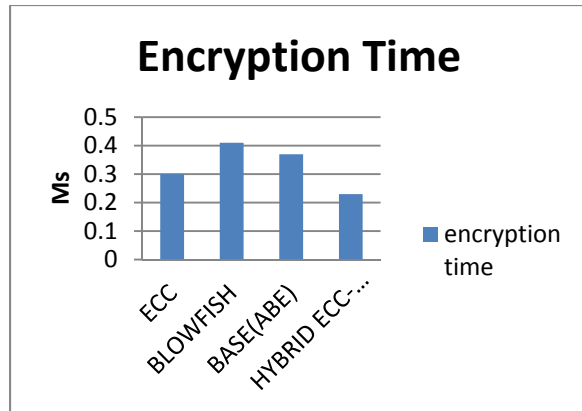


Figure 3 Encryption Time

Encryption time is used to estimate the speed of proposed system while working with the cloud users. The less time shows high speed communication between the user and cloud server. Here the system performs encryption time estimation of various existing approaches and proposed hybrid algorithm. The proposed architecture performs better in terms of encryption time as compared in the above figure.

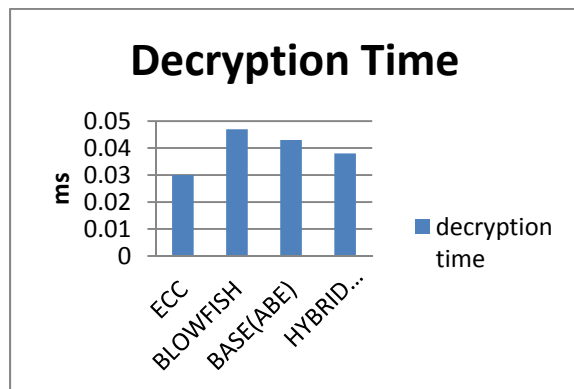


Figure 4 Decryption time

The next parameter is decryption also used to find the speed of the system. While user want to download or update the file, System find the file from repository and use access keys to decrypt and generate original format. The speed of decryption is also matter while user requests the files. The proposed hybrid algorithm performs better in terms of decryption time in above figure and compared with other techniques.

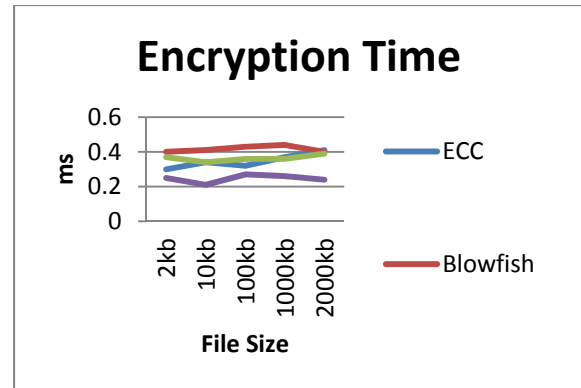


Figure 5 Encryption Time

Performance measurement is totally depends upon the test cases. The proposed architecture and various other algorithms performed on different file sizes for the encryption time measurement. In all cases proposed architecture perform better than all other existing approaches. With stable encryption time the system performance showed better in above figure.

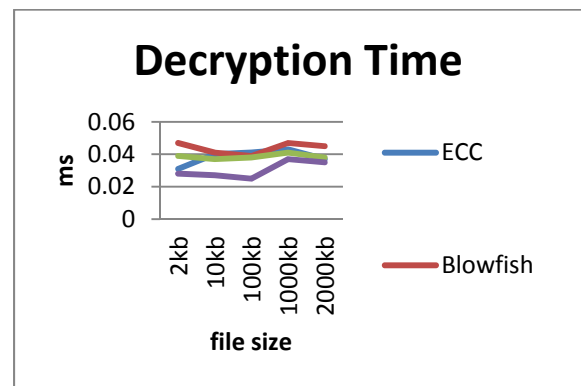


Figure 6 Decryption Time

Again Performance measurement in terms of decryption time is also depends upon the test cases. The proposed architecture and various other algorithms performed on different file sizes for the decryption time measurement. Here in all the cases proposed architecture perform better than all other existing approaches. With the stable decryption time the system performance showed better in above figure.

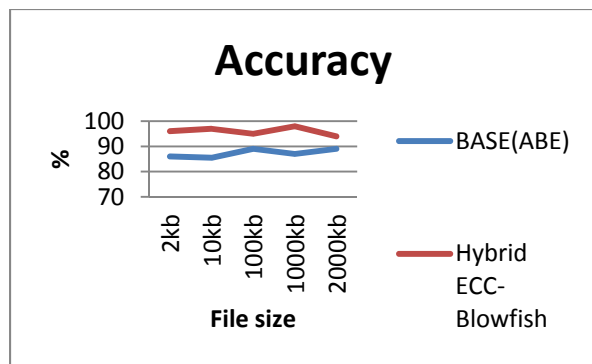


Figure 7 Accuracy

The accuracy factor is used to check the originality of decrypted content. While decrypting the encrypted file into original content, the error probability will be high if the algorithm is not efficient. Here in this system as compared with the existing system is performing well in all the cases.

VI. CONCLUSION AND FUTURE SCOPE

In this research work, conclusions have proposed an effective and efficient data access policies method for big data. Where the access the policies would not leak any privacy information. The access control choices are very significant for any shared network or system. However for a huge division system likes as a cloud network, access decision requires being more flexible and scalable. We have also implemented a hybrid approach (Blowfish and ECC encryption) method to calculate whether the data access policies. In order to enhance the efficiency, a new methods (Hybrid approach i.e, ECC and Blowfish) method has been implemented to discover the accurate the number of attributes in the matrix access. In this proposed work implemented new approach to show that scheme could preserve the privacy from any cloud access policies or services without employing high overhead. In future scope, it will focus on how to deal with the offline attributes /variables guessing intruders that identify the guessing string attributes by continually querying the ABF. It will implement a novel approach, to reduce the decryption time and probability factors to provide an improve and highly security in cloud storage.

REFERENCES

1. Sandhu, R. S., Coyne, E. J., Feinstein, H. L., & Youman, C. E. (1996). Role-based access control models. *Computer*, 29(2), 38-47.
2. Kuhn, D. R., Coyne, E. J., & Weil, T. R. (2010). Adding attributes to role-based access control. *Computer*, 43(6), 79-81.
3. Müller, H., Michoux, N., Bandon, D., & Geissbühler, A. (2004). A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. *International journal of medical informatics*, 73(1), 1-23.
4. El-triby, S., Mohamed, E. M., & Abdul-kader, H. S. (2012). Modern encryption techniques for cloud computing. In *ICCIT* (pp. 800-805).
5. Kumar, R., & Anil, A. (2011). Implementation of elliptical curve cryptography. *IJCSI International Journal of Computer Science Issues*, 8(4).
6. Arora, R., Parashar, A., & Transforming, C. C. I. (2013). Secure user data in cloud computing using encryption algorithms. *International journal of engineering research and applications*, 3(4), 1922-1926.
7. Yang, K., Han, Q., Li, H., Zheng, K., Su, Z., & Shen, X. (2017). An efficient and fine-grained big data access control scheme with privacy-preserving policy. *IEEE Internet of Things Journal*, 4(2), 563-571.
8. Yuan, Q., Ma, C., & Lin, J. (2015, January). Fine-grained access control for big data based on CP-ABE in cloud computing. In *International Conference of Young Computer Scientists, Engineers and Educators* (pp. 344-352). Springer, Berlin, Heidelberg.
9. Bandagar, V. S., & Kumbhar, H. V. (2015, December). Review Paper on Secure efficient data access control mechanism P2P storage cloud.
10. Nabeel, M., & Bertino, E. (2012). Privacy-Preserving Fine-Grained Access Control in Public Clouds. *IEEE Data Eng. Bull.*, 35(4), 21-30.
11. Zaghdoudi, B., Ayed, H. K. B., & Harizi, W. (2016, November). Generic Access Control System for Ad Hoc MCC and Fog Computing. In *International Conference on Cryptology and Network Security* (pp. 400-415). Springer, Cham.
12. Zhang, Y. Q., & Wang, X. Y. (2015). A new image encryption algorithm based on non-adjacent coupled map lattices. *Applied Soft Computing*, 26, 10-20.
13. Shabir, Muhammad Yasir, Asif Iqbal, Zahid Mahmood, and Ata Ullah Ghafoor. "Analysis of classical encryption techniques in cloud computing." *Tsinghua Science and Technology* 21, no. 1 (2016): 102-113. (15)
14. Bhanot, R., & Hans, R. (2015). A review and comparative analysis of various encryption algorithms. *International Journal of Security and Its Applications*, 9(4), 289-306

THE CHALLENGES OF E-GOVERNANCE IMPLEMENTATION IN NIGERIAN AVIATION INDUSTRY

BY

ELEBERI EBELE LETICIA

Department of Computer Science

Imo State University, PMB 2000, Owerri Nigeria

E-mail: ebyfav2001@yahoo.com

And

EKWOWUNNE EMMANUEL

Department of Computer Science

Imo State University, PMB 2000, Owerri Nigeria

ABSTRACT

In this study, the researcher evaluated the challenges of e-governance implementation in Nigerian aviation industry, using Dana Airline. The objectives of the study are; to examine the factors that hinder the effective implementation of e-governance in the selected airline in aviation industry in Nigeria; and to examine if the factors identified in the implementation of e-governance have significantly affected the performance of selected airlines in aviation industry in Nigeria. Recorded population of the study is 850, and the study used Yaro Yamane formula at 95% confidence level to obtain a sample size of 272. Cronbach alpha was employed to obtain a reliability instrument that yielded an index coefficient of 0.843, which made the instrument reliable. In line with the design of this study, the data that were collected for this study were analyzed using both descriptive and inferential statistics. The objectives posed for the study were answered using mean, standard deviation, and sample independent t-test statistics. The hypothesis was tested at 5% level of significance. Based on the findings of the study, it was concluded that ICT infrastructure is the most factor that hinders the implementation of e-governance in Nigerian aviation industry. This means that without proper ICT infrastructure, it is impossible to implement the e-governance in aviation industry in a developing country like Nigeria. In the absence of proper awareness among the users of the e-governance system, it is impossible to set up an effective e-governance system. Non-acceptability of IT systems, Low financial capability, Lack of electricity, High-cost, low-reliability of Internet access, Lack of training facilities and Lack of planning are all factors that hinder the effective implementation of e-governance in aviation industry in Nigeria. The study also concluded that the factors/challenges identified in the implementation of e-governance have significantly affected the performance of selected airlines in aviation industry in Nigeria. The study therefore recommends among others based on the findings of the study that it is paramount to have a proper ICT infrastructure to implement e-governance; government should take appropriate steps to enhance the awareness among the users of e-governance by organizing seminar meeting and other enlighten the users in the proper application of e-governance.

Keywords: Aviation Industry, E-governance, Information and Communication Technologies (ICT), Nigeria

Introduction

It is obvious that as Information technology is going, one day every organization will fully embrace information system of its own. Informatics is the science of information, and it studies the processing, representation, and communication of information technology in natural and artificial systems (Fourman, 2002).

A vast area of science with having information system as well as technology is known as Informatics. However, the present study focuses on e-governance which is closely connected to a great extent with the discipline of informatics. E-governance is a concept that is highly accepted by practitioners and scholars in the information and communication technology (ICT) domain. In this study, the e-governance implementation challenges are those for which e-governance application face obstacles or are deviated from reaching its expected target. Despite the fact that e-governance has numerous benefits, its challenges are also numerous in its implementation, especially in developing countries like Nigeria.

The Nigerian aviation industry has been frustrated with poor infrastructural facilities which constitute as threat to the management of the sector. According to Dode (2007), this has affected the performance of most airline industries in operation. It on record that most airline operators have withdrawn their services from Nigeria considering the un-conducive business domain encountered. Hence, if the challenges of e-governance implementation are identified and tackled, e-governance implementation in Nigerian aviation industries could go a long way in increasing revenue, promoting competitiveness and enhancing marketing in the public sectors. This study focuses on e-governance implementation challenges in Nigerian aviation industries, but the implementation has two ends; the two ends are the government (provider) and the staff (receivers). For the sake of this study, only the receivers (staff) of Nigerian aviation shall be considered to examine the challenges the face.

Meaning of E-Governance

It is necessary to understand the term governance before proceeding to e-governance. The word "governance" means the technique of decision-making and the approach by which decisions are executed (or not executed). The phrase "governance" can be employed in many circumstances like corporate governance, international governance, national governance and local governance. Governance can be seen as the compound techniques, processes, associations and institutions by which citizens and categories articulate their interests, exercise their rights and responsibilities and conciliate their disagreements (Olufemi, 2012)

It has been normally accepted that e-governance proffers enough future to enlarge the influence of government pursuits for citizens, which implies that the meaning of e-governance is wholly different and wide (Fang, 2002). The phrase e-governance simply means the application of information technologies like the Internet, World Wide Web, and mobile computing by government agencies that can change their association with citizens, businesses, various areas of government, and other governments. These technologies assist to carry out government services

to citizens, enhance interactions with businesses and industries, and provide entrance to information. The phrase e-governance can be explained as the application of emerging information and communication technologies to ease the procedures of government and public administration (Moon, 2002). E-governance according to Basu (2004) means the application by government agencies, like the aviation industry, of information technologies that have the capability to change relations with citizens, businesses and other arms of government.

Statement of problem

There are problems associated with e-governance implementation in aviation industries especially in developing countries, such as psychological as well as technical. It is imperative to accommodate good enough to the present-day situation of the e-governance application area to prevent bad user reactions. User acceptance is one of the most important quality factors of an e-governance. Implementing e-governance in developing countries like Nigeria is not rosy, as it has taken the first step towards applying e-governance and is encountering difficulties and will encounter challenges in future before achieving user acceptance. The aviation industry in developing countries is left behind as it faces a lot of challenges in implantation of e-governance. Hence, in this research, the main focus is to ascertain the challenges of e-governance implementation in Nigerian aviation industry and make it user friendly.

Aim and Objectives of the Study

The aim of this study is to evaluate the challenges of e-governance implementation in Nigerian aviation industry, using Dana Airline, the specific objectives are

- i. To examine the factors that hinder the effective implementation of e-governance in the selected airline in aviation industry in Nigeria
- ii. To examine if the factors identified in the implementation of e-governance have significantly affected the performance of selected airlines in aviation industry in Nigeria

Related Literature Review

Abasilim and Edet (2015) carried out a research on E-Governance and its implementation challenges in the Nigerian Public Service. In the study, the researchers said that E-governance is an improved tool that is geared in regards to effective public service delivery that is postulated on the expectation that the significant use of Information and Communication Technologies (ICT) technique in the day to day tasks of government will bring productive service delivery. It was as a result of many confrontations that hinder the effectual application of e-governance in Nigerian public service that led researcher to identifying some confrontations to e-governance application in Nigerian public service. The study did not employ any strong statistical analysis, as it was based on quality related study done by past researchers and inferences were drawn from

them, and the findings concluded that e-governance was the ultimate in encouraging transparency and accountability in government business. The study further recommended that government should be more committed to the application of e-governance, and also embarks on sufficient enlightenment about e-governance.

Okeudo and Nwokoro (2015) worked on Enhancing Airlines Operations through ICT Integration into Reservation Procedures: An Evaluation of Its Prospects in Nigeria. The study assessed the impact of ICT enhanced reservation procedures on the performance of airline industries with an intention that the information provided will guide airline operators and policy makers in their bid to sustain productivity and maintain efficiency. The study adopted an exploratory framework to evaluate the role of Airline Reservation System on the performance of airline companies with offices located in Sam-Mbakwe International Cargo Airport Owerri, Imo state Nigeria as the target populations. Two hypotheses were guided to achieve the objectives of the study, and the findings of the study revealed that there is significant relationship between the use of airline reservation system and the performance. Again there is correlation between the performance of an airline (Return on Asset) and the use of the Airline Reservation system.

Binuyo et al (2016) embarked a Study of the Application of Information and Communications Technology in Customer Relationship Management in Selected Airlines in Nigeria. The study examined the Customer Relationship Management (CRM) practices employed in selected airlines in the Nigerian Aviation industry. Again, the researchers conducted an enquiry on the factors affecting the successful deployment of Information and Communications Technology (ICT) for CRM and determined the effects of ICT on the performance of the industry. The study was carried out in the Head Offices of the local airlines (Lagos state and the Federal Capital Territory Abuja). The sampling technique employed was a multistage, which was used to choose ten local airlines and ten travel agencies. A random sample of two hundred Airline passengers was chosen for the study. The method of data collection was by Primary means via questionnaire. The data collected were collated and analyzed using statistical techniques such as descriptive and inferential statistics. The result of the analysis revealed that the adoption of ICT in airlines operations significantly reduced operational costs, improved service quality and improved identification of high value customers; hence concluded that the effective deployment of ICT assisted the Airlines in rendering better services to their passengers and ease an utmost performance of their operations.

Having reviewed these past researches, this study shall focus on the challenges of e-governance implementation Nigerian aviation industry.

Methodology

In line with the design of this study, the data that were collected for this study were analyzed using both descriptive and inferential statistics. The objectives posed for the study were answered using mean, standard deviation, and sample independent t-test statistics. The hypothesis was tested at 5% level of significance.

Sample Size Determination and Questionnaire Distributed

Onyenakeya (2001) states that sample are the number of people drawn from a population large and good enough to represent the entire population. A representative size is an essential requirement of any research study. As a result, it is pertinent to apply a mathematical approach to obtain such representative sample. However, the mathematical analysis on how the sample size will be derived is shown using Yaro Yamane Formula.

$$n = \frac{N}{1 + N(e)^2}$$

Where: n = sample size
 N = population size
 e = Allowable errors

There: $n = \frac{850}{1 + 850(0.05)^2}$

$$n = \frac{850}{1 + 850(0.0025)}$$

$$n = \frac{850}{1 + 2.125} = \frac{850}{3.125} = 272$$

Base on the calculation, the sample size is 272.

A total figure of two hundred and seventy two (272) was distributed in the selected airline (Dana Airline) to the respondents (airline agencies and airline officials) using purposive sampling technique. Out of the total figure distributed, two hundred and fifty (250) questionnaires were retrieved, that is 91.9%, while twenty two (22) questionnaires were not retrieved, which is 8.1%.

Reliability of the Instrument

The reliability of the instrument was achieved through a one-shot method of trial testing using thirty (30) respondents. The instruments were administered to the group and the scores were collated. Their responses (scores) were analyzed using Cronbach alpha which yielded an index coefficient of 0.843 via SPSS package as displayed in Table 1. The researcher therefore considered the instrument suitable and adequate for the study.

Table 1: SPSS output for the Reliability Test

Reliability Statistics

| Cronbach's Alpha | Cronbach's Alpha Based on Standardized Items | N of Items |
|------------------|--|------------|
| .843 | .846 | 8 |

Results and Discussion

Table 2: The following factors are the challenges that hinder the effective implementation of e-governance in Nigerian aviation industry.

| S/N | Indicators | \bar{X} | SD |
|-----|---|-------------|-------------|
| 1 | Lack of ICT infrastructure | 3.72 | 0.74 |
| 2 | Non-acceptability of IT systems | 3.53 | 0.72 |
| 3 | Lack of awareness | 3.54 | 0.70 |
| 4 | Low financial capability | 3.56 | 0.71 |
| 5 | Lack of electricity | 3.58 | 0.77 |
| 6 | High-cost, low-reliability of Internet access | 3.52 | 0.73 |
| 7 | Lack of training facilities | 3.57 | 0.72 |
| 8 | Lack of planning | 3.58 | 0.72 |
| | Cluster mean | 3.58 | 0.73 |

Key: VLE= Very Large Extent (4 Points), LE = Large Extent (3 Points), LE=Low Extent (2 Points) and VLE =Very Low Extent (1 Point)

From Table 2, all the factors considered in this study obtained an approximate average value of 4.00 which implies that they are the factors that hinder the effective implementation of e-governance in Nigerian aviation industry to a very large extent.

Table 4: SPSS Output of factors identified in the implementation of e-governance on Airline Performance

| One-Sample Statistics | | | | | | |
|-----------------------|---|--------|----------------|-----------------|--|--|
| | N | Mean | Std. Deviation | Std. Error Mean | | |
| VAR00001 | 8 | 3.5750 | .06279 | .02220 | | |

| One-Sample Test | | | | | | |
|-----------------|----------------|----|-----------------|-----------------|---|--------|
| | Test Value = 0 | | | | | |
| | t | df | Sig. (2-tailed) | Mean Difference | 95% Confidence Interval of the Difference | |
| | | | | | Lower | Upper |
| VAR00001 | 161.033 | 7 | .000 | 3.57500 | 3.5225 | 3.6275 |

From the SPSS output, the p-value (0.000) is less than 0.05, which implies that the factors/challenges identified in the implementation of e-governance have significantly affected the performance of selected airlines in aviation industry in Nigeria.

Conclusion

Having conducted the analysis, it has been concluded that ICT infrastructure is the most factor that hinders the implementation of e-governance in Nigerian aviation industry. This means that without proper ICT infrastructure, it is impossible to implement the e-governance in aviation industry in a developing country like Nigeria. In the absence of proper awareness among the users of the e-governance system, it is impossible to set up an effective e-governance system. There should be adequate knowledge of the users of e-governance system and its services. Non-acceptability of IT systems, Low financial capability, Lack of electricity, High-cost, low-reliability of Internet access, Lack of training facilities and Lack of planning are all factors that hinder the effective implementation of e-governance in aviation industry in Nigeria. The study also concluded that the factors/challenges identified in the implementation of e-governance have significantly affected the performance of selected airlines in aviation industry in Nigeria.

The study therefore recommends among others based on the findings of the study that it is paramount to have a proper ICT infrastructure to implement e-governance; government should take appropriate steps to enhance the awareness among the users of e-governance by organizing seminar meeting and other enlighten the users in the proper application of e-governance.

References

- Dode, R.O. (2007). Prospects of E-Government Implementation in Nigeria. 1st International Conference on Theory and Practice of Electronic Governance, Macao, China, pp. 380-383, December 10-13.
- Fourman, M. (2002). Informatics Research Report EDI-INF-RR-0139
- Olufemi, F.J. (2012). Electronic Governance: Myth or Opportunity for Nigerian Public Administration? *International Journal of Academic Research in Business and Social Sciences* September 2012, Vol. 2, No. 9
- Moon, M. J. (2002). "The Evolution of E-governance Among Municipalities: Rhetoric or Reality?" *Public Administration Review*, 62(4).
- Basu, S, (2004). "E-governance and Developing Countries: An Overview", *International Review of Law Computers*, 18(1)
- Abasilim, U.D. and Edet, L.I. (2015). E-Governance and Its Implementation Challenges in the Nigerian Public Service. *Acta Universitatis Danubius. Administratio*, Vol 7, No 1 (2015).
- Fang, Z. (2002), E-government in Digital Era: Concept, Practice and Development. *International Journal of the Computer, the Internet and Management*, Vol.10, No. 2.
- Okeudo, J.N. and Nwokoro, I.A. (2015). Enhancing Airlines Operations through ICT Integration into Reservation Procedures: An Evaluation of Its Prospects in Nigeria. *British Journal of Economics, Management & Trade* 8(3): 190-199, 2015, Article no.BJEMT.2015.110
- Binuyo, G.O., Olasupo, J.O., Ogunjemilua, E.M. (2016). A Study of the Application of Information and Communications Technology in Customer Relationship Management in Selected Airlines in Nigeria. *International Journal of Computer Applications* (0975 – 8887) Volume 139 – No.1, April 2016

Ensemble Feature Selection(EFS) And Ensemble Convolutional Neural Network With Vote Classification (ECNN-VC) Approach For Detection Of Snore Sounds In Obstructive Sleep Apnea Diagnosis

E.Jeslin Renjith,
Research Scholar,
Bharathiar University, Coimbatore,India
Jeslin.renjith@gmail.com

Dr.Christy.A,
Research Supervisor,
Bharathiar University,Coimbatore,India
ac.christy@gmail.com

Abstract— Examining deviations in the upper airway reduction for the period of sleep is invasive and expensive. Because the snoring sounds are created by air instability and vibrations of the upper airway appropriate to its reduction. These sounds have been used as a non-invasive method to evaluate upper airway narrowing. Snoring is able to be a sign of Obstructive Sleep Apnea (OSA) disease. Recently an Ensemble Swarm Intelligent based VOTE Classification (ESIVC) approach is introduced for snore sound detection and diagnosis of OSA. But in ESIVC depends on single feature selection algorithm which mightn't provide higher performance. To solve this problem in this work we introduce an Ensemble Feature Selection (EFS) by combining many feature selection methods. The major aim of this work is to introduce a subject-precise acoustic model of the upper airway in the direction of examine the impact of upper airway by snoring sounds of multi-features. The initial stage of the work the collected audio signals recorded using Drug-Induced Sleep Endoscopy (DISE) and these signals were digitized by the frequency rate of 44.1 kHz, Pulse Code Modulation (PCM), and down-sampled to 16 kHz per each sample, which is the lowest sampling rate of the audio recorder. Secondly the noises presented in the audio signals were removed by Wiener-Filter (WF) algorithm. Some of the features like Crest Factor, original Frequency, Spectral Frequency Features, Subband Energy Ratio, Mel-Scale Frequency Cepstral Coefficients (MFCC), Empirical Mode Decomposition (EMD) Features, and Wavelet Energy Features have been extracted from the noise suppressed audio signals and input into EFS approach. Then extracted features are selected by the use of EFS, where multiple FS methods are integrated to provide more detection results in individual subjects with higher airway. EFS algorithm fusion the results of all FS methods in combination with an Ensemble Convolutional Neural Network with VOTE Classification (ECNN-SVC) approach showed the best classification results by subject independent validation. The idea behind ECNN-SVC approach is to generate a first CNN which is coarsely optimized however gives an excellent opening pointing for further tuning, which will serve as the ECNN approach. Then, the obtained weights are fine tuned by using back propagation several times to create an ensemble of CNNs

with the purpose of representing the snoring sounds. ECNN-VC approach is help to classify the snore sounds in the upper airway. From the detection results, we found that proposed ECNN-VC classifier significantly performs better in the upper airway. These results encourage the use of snoring sounds examination to evaluate the upper airway analysis during OSA.

Keywords— Obstructive Sleep Apnea (OSA); Ensemble Feature Selection (EFS), Ensemble Convolutional Neural Network with VOTE Classification (ECNN-VC) , Velum Oropharyngeal Tongue Epiglottis (VOTE), Bat Algorithm (BA).

1. INTRODUCTION

Obstructive Sleep Apnea (OSA) syndrome is a most dangerous sleep disorder described by means of the frequent closing of the upper airway for the duration of sleep, between adults 30–70 yrs of age, roughly 13% of men and 6% of women have reasonable in the direction of severe OSA, 14% of men and 5% of women have an Apnea-Hypopnea Index (AHI) ≥ 5 plus symptom of morning sleepiness [1]. It is moreover being predictable as a self-determining risk factor used for many clinical consequences, together with daytime sleepiness [2], total hypertension, and improved risk of cardiovascular and cerebrovascular syndrome [3-4], traffic accidents, [5] and impaired value of life [6].

Polysomnography (PSG) is presently the best of OSA diagnosis [7]. But, PSG needs a full-night hospital continue in a particularly operational sleep set, coupled to more than 15 channels of dimensions needing physical contact by means of sensors [8]. So PSG is difficult, costly, and not appropriate for collection screening. The restricted number of PSG facilities approximately the world has long waiting lists, representing it impractical to test each and every one the patients in need of such measurement. Roughly 80–90% of patients by means of OSA are supposed to be undiagnosed [9]. With progress in tools and the improvement of manageable monitors, home testing designed for sleep associated breathing disorders is currently

sufficient and avoid several of the issues of a focused in-laboratory polysomnogram.

Though the manageable monitors might advance the analysis of OSA for several, it is important with the purpose of health care professionals by these algorithms which recognize several inherent issues. There is a huge necessitate for a basic screening instrument able of suitable and consistent analysis of OSA. Loud snoring, as a characteristic sign of OSA, is reported in more than 80% of OSA patients. The acoustic characteristics of snoring have been examined by several authors in acoustics and otorhinolaryngology with the goal of introducing a classification method for the detection of OSA, PSG [10]. The work done based on the classifiers for snoring sound examination provides higher classification accuracies up to 80% in the classification of OSA [11].

But in general the snoring sounds have been recorded suitably with a microphone positioned on the neck, or in the neighborhood of the patient in the room. Snoring has been defined as the symptom of OSA [12]; a frequent respiratory disorder current in around 10% of adult people. OSA is differentiating by means of repetitive complete or incomplete collapse of the upper airway for the duration of sleep. Accordingly, several number of works on acoustic analysis of snoring sounds paying attention on the relationship among OSA severity and a selection of snoring sounds features like intensity [13], power spectral features, bi-spectral and non-linear measures [14], formant frequencies [15] and temporal features [16]. Recently many of the works have been performed for OSA detection systems [17] and AHI estimation depending on entire-night audio recording of snoring [18-19]. For devices make use of acoustic signals, the information are inadequate to validate whether the make use of acoustic signals with other signals as a vary designed for airflow is sufficient to identify OSA [20]. On the other hand, only less analysis was placed on the sound effects of deviation in the upper airway anatomy for the duration of sleep on snoring sounds features.

Only restricted number of works has addressed the problem of automatic detection for snore signals, and less number of works regarding snore detection using upper airway signals. Several snore/non-snore detection methods have been proposed to examine snore sound events. But all of the works has majorly focuses on analyzing the assured well-chosen acoustic features are considered for their emotion to the anatomical systems of snoring sound detection. Furthermore, sequentially in the direction of continuously approximation of severity and unpredictability of an individual's snore, the video recording of a complete night is necessary. The major objective of this paper is to introduce and experiment an Ensemble Feature Selection (EFS) and Ensemble Convolutional Neural Network with VOTE Classification (ECNN-VC) approach with high efficiency. The function of EFS for multi-feature analysis has been proposed in this work [21]; however, advanced feature extraction, and classification approach have not yet been used for this purpose. In this work EFS approach, we combine the results of many FS algorithms, and introduce newly selected features within advanced ECNN-VC approach. The proposed ECNN-SVC approach showed the better detection results for snore sounds of 40 male patients

has been recorded by using DISE, and classified by using Ear, Nose & Throat (ENT) experts.

2. LITERATURE REVIEW

Alencar et al [16] find with the purpose of the many of asymmetrical snores which is straightforwardly available and quantified by what we describe the Snore Time Interval Index (STII). It is in high-quality agreement with the recognized AHI, which states the severity of OSA and is founded only from PSG. Additionally, the Hurst examination of the snore sound itself, which determines the fluctuations in the signal as a function of time period, is used in the direction of construct learning with the purpose of able to differentiate among patients with no or mild AHI and patients with moderate or severe AHI.

Jin et al [22] proposed a new method to analytically establish the performance of acoustic analysis of snoring in the diagnosis of OSA by means of a meta-analysis. The results are measured using the metrics like sensitivity, specificity, and accuracy for acoustic examination of snoring in the analysis of OSA. The medium of AHI threshold was 10 events/h, and the range of 5-15 is suitable to this work. The results demonstrated with the purpose of the mutual estimates were constant and consistent.

Xu et al [23] proposed a metabolomics approach in the direction of evaluate urinary metabolites in three different types of members: patients among PSG-confirmed OSA, Simple Snorers (SS), and usual subjects. Ultra-reults liquid chromatography together with quadrupole time-of-flight mass spectrometry and gas chromatography together with time-of-flight mass spectrometry were used in this work. Metabolic pathways connected with SS and OSA was well-known by means of the metabolomics system, and the changed metabolite signatures might potentially serve as a different diagnostic system to PSG.

Karunajeewa et al [24] proposed a new method for detecting Obstructive Sleep Apnea Hypopnea Syndrome (OSAHS) depending on the snore sound analysis. The proposed method introduces a logistic regression model fed with snore constraints extracted from its features like pitch and Total Airway Response (TAR) determined by means of using a Higher Order Statistics (HOS) features. The results demonstrated that the detection of snore-depending on OSAHS device shouldn't need some contact measurements.

Saha et al [25] proposed to develop a new subject-exact acoustic algorithm in order to analysis the importance of upper airway anatomy by means of snoring sounds features. The results of this subject-exact acoustic algorithm were evaluated using the 20 male individuals of snoring sounds with their features like intensity and resonant frequencies at the period of sleep. It concludes that the proposed work performs better and encourages or motivates to other researchers to use snoring sounds examination to evaluate the upper airway anatomy for the period of sleep.

Abeyratne et al [26] introducing a new snore based multi-feature class OSA screening system by means of combining features of snore sounds. Snore sound feature classes of individuals are optimized using logistic regression for improving detection results. Accordingly, each and every

one feature classes were combined and optimized to obtain better detection results.

Praydas et al [27] proposed a new method to differentiate the severity of OSA patients. The proposed algorithm uses a K-Means clustering to group the Sound Spectrum and rebuild new features. Then multiclass classification is performed by using a Support Vector Machine (SVM) for the detection of snore sounds related to their severity. From the results it concludes that the proposed system achieved 75.76% of accuracy and it is able to give higher investigative suggestions designed for OSA screening.

Qian et al [28] proposed a new multi-feature analysis in the direction of analytically the results of different acoustic features, and classifiers designed for their results in the detection of the excitation position of snore sounds. Then some of the features like crest factor, fundamental frequency, spectral frequency features, subband energy ratio, mel-scale frequency cepstral coefficients, empirical mode decomposition-based features, and wavelet energy has been extracted from the snore sound signals and given as input to feature selection algorithm. ReliefF is proposed for ranking of features and the ranked features have been evaluated with the classifiers. From the results it concludes that this approach provides better results when compared to other methods by considering multi-feature of snore sound creation in individual subjects.

Dedhia and Weaver [29] learn the test with subsequent relations: (1) Entire obstruction on DISE and PSG and individual actions of OSA; (2) language base and AHI. Every DISE video was evaluated designed for entire obstruction at Velum, Oropharynx, Tongue, Epiglottis (VOTE classifiers). Student's t test, correlation, and multivariate linear regression were performed to measure the dataset. From the study it concludes that the thorough examination is required to establish the significance of every site and degree of obstacle seen on DISE. All the above mentioned review work majorly focuses on analyzing the assured well-chosen acoustic features are considered for their emotion to the anatomical systems of snoring sound detection. Furthermore, sequentially in the direction of continuously approximation of severity and unpredictability of an individual's snore, the video recording of a complete night is necessary. Thus, developing an automatic snore detection algorithm with multi-features analysis to estimate complete night recordings in an appropriate and accurate manner might be useful which is focused in this work.

3. PROPOSED WORK

The major objective of this paper is to introduce an Ensemble Feature Selection (EFS) algorithm and experiment an Ensemble Convolutional Neural Network with VOTE Classification (ECNN-VC) approach with high efficiency, and perceptive whole-night snore sound detector based on individual subjects. The initial stage of the work the collected audio signals recorded using Drug-Induced Sleep Endoscopy(DISE) and these signals were digitized by the frequency rate of 44.1 kHz, Pulse Code Modulation (PCM), and down-sampled to 16 kHz per each sample, which is the

lowest sampling rate of the audio recorder. Secondly the noises presented in the audio signals were removed by Wiener-Filter (WF) algorithm. Some of the features like Crest Factor, original Frequency, Spectral Frequency Features, Subband Energy Ratio, Mel-Scale Frequency Cepstral Coefficients (MFCC), Empirical Mode Decomposition (EMD) Features, and Wavelet Energy Features have been extracted from the noise suppressed audio signals and input into EFS approach. Then extracted features are selected by the use of EFS algorithm which fusion the results of all FS methods. These selected features are classified using ECNN-VC approach which shows better classification results by subject independent validation. The proposed ECNN-VC approach showed the better classification results snore sounds of 40 male patients has been recorded by using DISE, and classified by using ENT experts depending on the VOTE classification [30]. ECNN-VC approach is help to classify the snore sounds in the upper airway. VOTE classifier which classifies the recordings into four major classes: the level of the Velum (V), the Oropharyngeal area consists of the palatine tonsils (O), the Tongue base (T), and the Epiglottis (E). During samples collection stage, Snoring sounds (SnS) with many vibration location or unknown base of vibration were removed from original records. From each integrated recording, three to five SnS which demonstrated no obstructive character, have been manually chosen. From the 40 subjects, 11, 11, 8, and 10 subjects were classified into four major classes. Among one and five snoring events per class were extracted per each person. On the whole for implementation work we have 164 snoring events (41 episodes for each sensor category of SnS, length ranging from 0.728 to 2.495s with an average of 1.498 s). They segmented the episodes into distinct parts for additional feature extraction, EFS and ECNN-VC approach. Each segment has duration of 200 ms and neighbouring segments have been extending beyond of 50 %. The overall architecture of the proposed EFS and ECNN-VC approach is illustrated in Figure 1.

A. Pre-processing

For design and validations phases, the collected audio signals recorded in the DISE were digitized by the frequency rate of 44.1 kHz, Pulse Code Modulation (PCM), and down-sampled to 16 kHz per each sample, which is the lowest sampling rate of the audio recorder. These signals were noise removed based on the adaptive noise suppression which follows the procedure of Wiener-filter (WF). This WF depends on automatically tracking background noise subdivision in order to estimate their spectrum and subtracting them from the audio signal. In this work, a noise spectral algorithm was subtracted from each audio frame (40 ms). Each frame's frequency part was suppressed by a particular suppression noise spectral algorithm, and it was restricted to the range [0, 225 dB] so as to prevent a key distortion when low SNR.

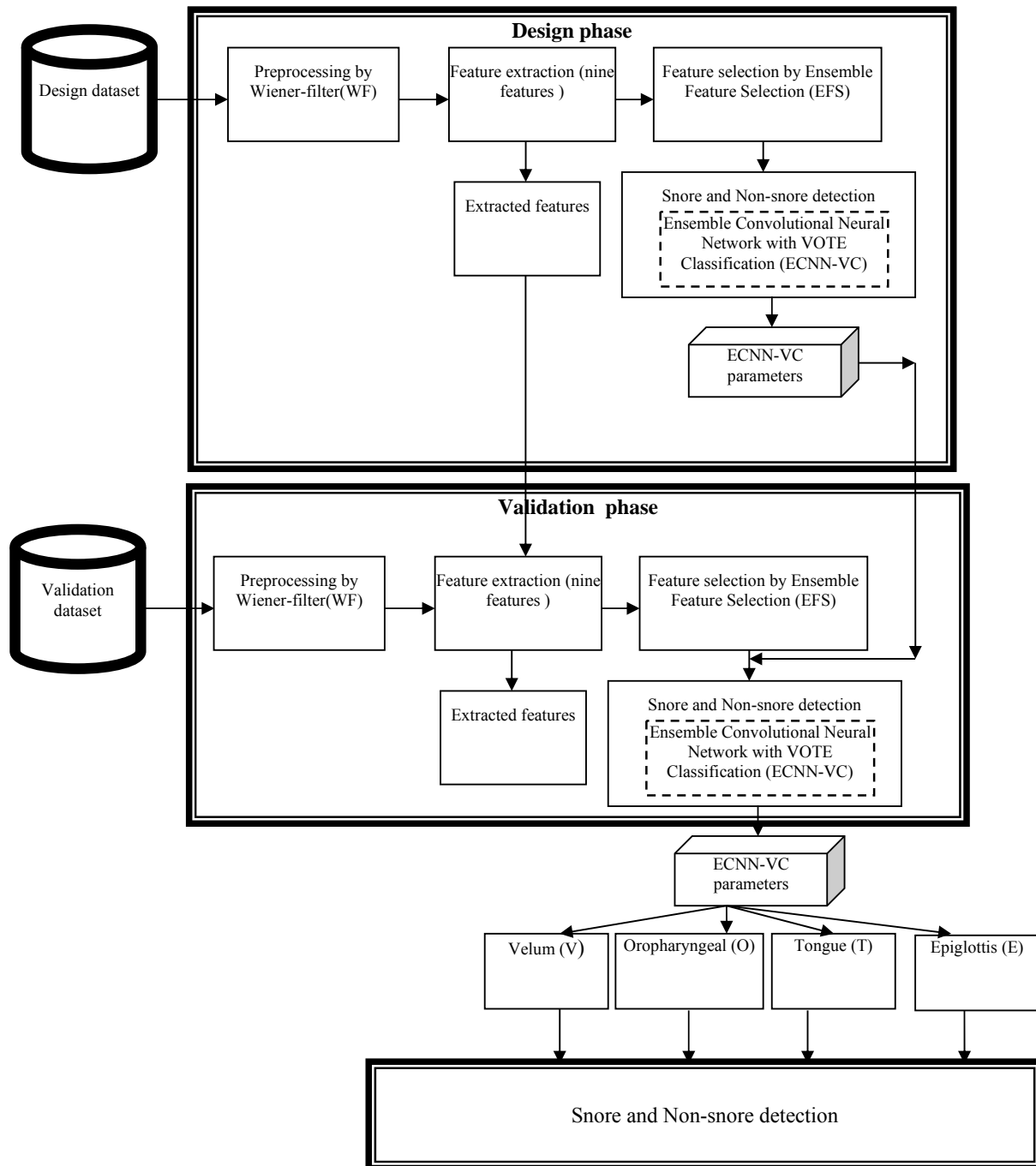


Figure 1. Architecture diagram

B. Ensemble Feature Selection(EFS)

In ensemble learning, is performed by a collection of several classifiers or algorithms is trained, and the final results of the ensemble is determined by combining the outputs of the single classifiers or algorithms, e.g. by means of using majority voting in the case of classification. In the recent work [37] concludes that the ensemble might performs better when compared to single algorithms when weak (unstable) algorithms are combined, generally since of three major reasons: a) many diverse however regularly optimal hypotheses be able to exist and the ensemble decreases the risk of selecting a incorrect hypothesis, b)

classifiers might end up in varied local optima, and the ensemble might provide a better estimate of the correct function, and c) the correct function shouldn't be characterized by some of the hypotheses in the hypothesis space of the classifier and by combining the results of the single algorithms, the hypothesis space might be extended.

Related in the direction of the case of supervised algorithm, ensemble algorithms should be used to increase the robustness of feature selection algorithms. Definitely, in huge feature/ little example size domain it is frequently statement with the purpose of many feature subsets might yield regularly optimal results [38], and EFS might decrease

the risk of selecting an uneven subset from the snore sounds. Furthermore, different feature selection algorithms may yield feature subsets that can be considered local optima in the space of feature subsets, and EFS might give a better approximation to the optimal subset or ranking of features. Finally, the representational power of a particular feature selector might constrain its search space such that optimal subsets cannot be reached. EFS could help in alleviating this problem by aggregating the outputs of several feature selectors.

There are two major steps in generating EFS. The initial step includes generating a set of many feature selection algorithms, every giving their output, whereas the second step combines the results of the single feature selection algorithms. Combining the results of many FS algorithms has been performed by using weighted voting, e.g. in the case of obtaining a consensus feature ranking, or by including the mainly commonly chosen snore features in the case of obtaining a final snore features subset.

In this work, we focus on EFS algorithm with the purpose of work by combining the features presented by the single feature selection algorithms into an absolute consensus ranking. Let us consider that the ensemble 'E' including of 's' feature selection algorithms, $E = \{Fe_1, Fe_2, \dots, Fe_s\}$, then we consider that every Fe_i gives a feature ranking $fe_i = (fe_i^1, \dots, fe_i^N)$, which are combined into a consensus snore features ranking 'fe' by means of weighted voting:

$$fe^l = \sum_{i=1}^s w(fe_i^l) \quad (4)$$

where $w(\cdot)$ denotes a weighting function. If a *linear aggregation* is carryout using $(fe_i^l) = fe^l$, this results in a sum where snore features give in a linear manner concerning their rank. By changing $w(fe_i^l)$, extra or fewer weight be able to be situate in the direction of the rank of each snore feature. This has been used to contain designed for rankings where top snore features be able in the direction of be required to influence the ranking considerably more than lower ranked snore features. The following feature selection algorithms are used in this work for snore features.

Filter algorithms be able to be measured as a pre-processing stage since they are independent on the ECNN-VC approach. The snore sound features subset is generated by measuring the connection between input snore sound features and output detection [39] of the current system. Also the snore sound features are sorted depending on their relevancy to the objective by calculating statistical tests. Fisher criterion [40] is one of the majorly used filter feature selection algorithm. The Fisher index $Fe(i)$ of the i th snore sound features is determined through (5) [41]:

$$Fe(i) = \frac{|\mu_1(i) - \mu_0(i)|}{\sqrt{\sigma_1^2(i) + \sigma_0^2(i)}} \quad (5)$$

where $\mu_j(i)$ and $\sigma_j(i)$ is denoted as the mean and standard deviation of the i -th snore sound feature in the two major

classes (0/1). This index highlights the significance of every feature determining the ratio among the variation between the means of the distribution of the two classes and the sum of their variance [42]. T-test is also considered as another most important filter approach and it used to determine the significance of snore sound features [43]. It obtains from a statistical analysis frequently used (t-test) and its index are determined as in equation (6):

$$t(i) = \frac{|\mu_1(i) - \mu_0(i)|}{\sqrt{\frac{\sigma_1^2(i)}{n_1} + \frac{\sigma_0^2(i)}{n_0}}} \quad (6)$$

where n_0 and n_1 are the number of snore sound features in the null and unitary class correspondingly.

Kullback Liebler distance (KL-distance) is a computed by measuring a probability distribution p in the direction of a target probability distribution q . Designed for discrete distributions $p=\{p_1, p_2, \dots, p_n\}$ and $q=\{q_1, q_2, \dots, q_n\}$ the KL-distance is determined as in equation (7).

$$KL(p, q) = \sum_i p_i \log_2 \left(\frac{p_i}{q_i} \right) \quad (7)$$

Wrapper feature selection algorithm treats the ECNN-VC approach as a black box making it common. The selection of the snore sound features has the main improvement of considering the ECNN-VC approach, on the other hand this ECNN-VC approach provides improved detection results than filter approach and it is not suitable for handling with huge snore sound features. A frequently used wrapper approach is the so named Greedy Search strategy which increasingly creates the snore sound features set by considering or removing single snore sound features from an initial snore sound features subset. Greedy search has been divided into two types: Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS).

SFS algorithm starts with an empty set of snore sound features and the remaining snore sound features are iteratively added until a fixed stopping criterion is achieved. Usually, related to OSA detection performance, the used criterion is the accuracy of the ECNN-VC approach performed. SBS is the opposite of the SFS, it starts together with all snore sound features and then the less important snore sound features are eliminated one by one. A snore sound feature is defined interesting and also left inside the set, if removing it the results of the ECNN-VC approach decreases.

C. Ensemble Convolutional Neural Network with VOTE Classification (ECNN-VC)

EFS algorithm fusion the results of all FS methods in combination with an Ensemble Convolutional Neural Network with VOTE Classification (ECNN-VC) approach showed the best classification results by subject independent validation. The idea behind ECNN-VC approach is to

generate a first CNN which is coarsely optimized however gives an excellent opening pointing for further tuning, which will serve as the ECNN approach. VOTE classifier which differentiate four levels inside the upper airway: the level of the Velum (V), the Oropharyngeal area consists of the palatine tonsils (O), the Tongue base (T), and the Epiglottis (E). Here the VOTE classifiers follows the procedure of ECNN approach which classifies the recordings into four major classes (V,O, T and E).

Ensemble Model

Let us consider that the $y_k^n(j)$ is denoted as the value of the k th output layer unit of the j^{th} CNN model for n^{th} input snore sound features. The linear and log-linear ensemble approach detection results for the same input snore sound features would be described as in equation (8-9)

$$En_{linear}^n = \arg \max_k \sum_{j=1}^l y_k^n(j) \quad (8)$$

$$En_{log_linear}^n = \arg \max_k \prod_{j=1}^l y_k^n(j) \quad (9)$$

where l is the number of Convolutional Neural Networks (CNNs) classifiers is aggregated to create final Ensemble classifier. Each CNNs model is chosen depending on the 5-fold cross validation test. In each network ECNN architecture and training organization, we combined the top-5 introduced single CNNs classifier to produce final Ensemble classifier by equations (8-9) of their outputs. In equation (8-9), CNNs is performed depending on the multilayer learning architecture which consists of the following layers input layer, convolutional layers, pooling layers, fully connected layers and the output layer. The major objective of this CNN is to study the hierarchy of fundamental snore sound features representations. In the following section explain the details of layers in CNN [44-45]:

Convolutional layer

At each convolutional layer three dimensional matrices (kernels) are decreased over the snore sound input features and set the dot product of kernel weights with the accessible field of the snore sound features as the resultant OSA detection results. This layer helps to maintain the comparative position of snore sound features to each other. The multi-kernel feature of convolutional layers permits them to prospectively mine several distinct snore sound features maps from the same input sounds.

Activation layer

The results from the convolutional layer are given as input to the activation function in the direction of correcting the negative results. Moreover we used the Rectified Linear Unit (ReLU) which is generally the chosen selection since of its ease, convergence, decreased likelihood of vanishing gradients and affinity in the

direction of adding sparsity over other sigmoid function. The output of j^{th} ReLU layer is specified its input was determined by using the following equation (10):

$$a_j^{out} = \max(a_j^{in}, 0) \quad (10)$$

Normalization layer

From the results of ReLU layer, then a Local Response Normalization (LRN) map is functional following the primary convolutional layers. These layers reduce the local ReLU neurons' activations because it is not required to limit them (Eq. 10). LRN [46], the limited regions are extended across neighbor snore sound feature maps at every spatial position. The results of j^{th} LRN layer is given as input to next layer and it is determined as in equation (11):

$$a_j^{out} = \frac{a_j^{in}}{\left(1 + \frac{\alpha}{L} \sum_{n=1}^L a_j^{in}(n)^2\right)^\beta} \quad (11)$$

where a_j^{in} is denoted as the n^{th} element of the snore sound features and L is the length of a_j^{in} feature vector. α , β and L are the layer's hyperparameters and are predefined as default values taken from [47] ($\alpha = 1$, $\beta = 0.75$ and $L = 5$).

Pooling layer

After rectification layer results are found then pooling layer is performed by considering some pooling operations. It builds up the values in a smaller region by means of subsampling functions like max, min, and average sampling. In this research work, max pooling operation is carryout in CNNs.

Fully connected Layer

From the results of the convolutional and pooling layers, the network creative layers are fully associated. These layers are normally consideration as the model of CNN classifier, since they consider the snore sound features which is extracted from convolutional layers and provides the detection results, which is the classifier detection.

Dropout Layer

From the following of fully connected layers then dropout layer are called which is considered as the final layer of the CNN classifier which produces the class-specific probabilities. In these layers, a subset of input neurons as well as all their connections is provisionally removed from the entire CNNs model.

Learning

Finally the training and testing of the CNN is performed via the use of Stochastic Gradient Descent (SGD) with two major steps: Forward and Back Propagation. In forward stage, the classifier formulates detections using the snore sound features in the training batch and the current classifier parameters. Once the

detection for all snore sound features is made, the loss is determined using the truth label provided by the ENT experts. Here use a softmax loss function which is computed as follows in equation (12):

$$L(t, y) = -\frac{1}{N} \sum_{n=1}^N \sum_{k=1}^C t_k^n \log \left(\frac{e^{y_k^n}}{\sum_{m=1}^C e^{y_m^n}} \right) \quad (12)$$

where t_k^n is the n^{th} training snore sound features k^{th} output class, and y_k^n is the value of the k^{th} output layer unit in response to the n^{th} input training snore sound features. N is the number of training snore sound features in the minibatch, and two class labels ($C = 2$). All the way through the back propagation, the loss gradient concerning to every classifier weights support improvement the weights as described as follows in equation (13):

$$W(j, i + 1) = W(j, i) + \mu \Delta W(j, i) - \alpha(j, i) \frac{\partial L}{\partial W(j)} \quad (13)$$

where $W(j, i)$, $W(j, i + 1)$ and $\Delta W(j, i)$ is denoted as the weights of j^{th} convolutional layer at iteration i and $i + 1$ and the updating of weight of iteration i , μ is the momentum and $\alpha(j, i)$ is the detection rate and is dynamically lowered as the training purpose. In order to evaluate the snore detection rate of all classifiers, apply the Unweighted Average Recall (UAR).

4. EXPERIMENTATION RESULTS

In this work, we analytically evaluate normally used multi-acoustic features for their detection accuracy on the classifiers of snore sounds depending on Logistic Regression (LR), k Nearest Neighbour (kNN), Ensemble Swarm Intelligent based VOTE Classification (ESIVC) and Ensemble Convolutional Neural Network with VOTE Classification (ECNN-VC) approach. All experiments are implemented with the help of MATLAB R2012 software environment.

During samples collection stage, Snoring sounds (SnS) with many vibration location or unknown base of vibration were removed from original records. These SnS records have been then extracted from the audio signals and classified based on the ECNN- VC approach. From the 40 subjects, 11, 11, 8, and 10 subjects were classified into four major classes. On the whole for implementation work we have 164 snoring events (41 episodes for each sensor category of SnS, length ranging from 0.728 to 2.495s with an average of 1.498 s). They segmented the episodes into distinct parts for additional feature extraction, EFS and ECNN-VC approach. In order to evaluate the snore detection rate of all classifiers, apply the Unweighted Average Recall (UAR), described as follows:

$$UAR = \frac{\sum_{class=1}^{N_{MC}} N_{class,correct} / N_{class,all}}{N_{MC}} \times 100\% \quad (14)$$

D. Without feature selection

The UAR results of three classifiers with different feature sets are shown in Table I.

Table 1. UAR ([%]) Results Obtained With Nine Features And Four Classifiers Without Feature Selection

| Features | ECNN-VC | ESIVC | LR | k-NN |
|---------------------|---------|-------|-------|-------|
| Crest Factor | 52 | 48 | 39 | 36 |
| F0 | 58 | 43 | 38 | 35 |
| Formants | 65 | 62 | 56 | 54 |
| SFF | 75 | 72 | 67 | 63 |
| PR | 53 | 48 | 43 | 38 |
| SER | 83 | 75 | 68 | 65 |
| MFCCs | 86 | 82 | 79 | 76 |
| EMDF | 70 | 61 | 53 | 47 |
| WEF | 82 | 78 | 65 | 59 |
| ALL | 91 | 87 | 75 | 63 |
| Average | 69.33 | 63.22 | 56.44 | 52.55 |

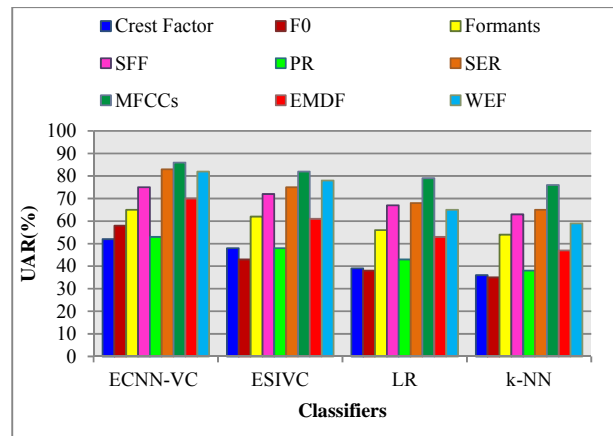


Figure 2. UAR results vs. four classifiers without feature selection

From the results it concludes that the proposed ECNN-VC approach with the purpose of MFCCs achieves the best snore sound detection results of 86%. Among them nine of the different subsets, the novel SER performs provides next higher UAR of 83% when compared to other classifiers namely Logistic Regression (LR), k Nearest Neighbour

(kNN) classifier and Support Vector Machine (SVM) classifier. On the other hand, Crest Factor, F0, and PR800 doesn't give better detection results for all classifiers is illustrated in figure 2. From the results it concludes that the proposed ECNN-VC approach produces average accuracy results of 69.33% which is 6.11%, 12.88% and 16.77% higher when compared to ESIVC, LR, and kNN classifiers respectively.

Table 2. Error rate results obtained with nine features and four classifiers without feature selection

| Features | ECNN-VC | ESIVC | LR | k-NN |
|---------------------|---------|-------|-------|-------|
| Crest Factor | 48 | 52 | 61 | 64 |
| F0 | 42 | 57 | 62 | 65 |
| Formants | 35 | 38 | 44 | 46 |
| SFF | 25 | 28 | 33 | 37 |
| PR | 47 | 52 | 57 | 62 |
| SER | 17 | 25 | 32 | 35 |
| MFCCs | 14 | 18 | 21 | 24 |
| EMDF | 30 | 39 | 47 | 53 |
| WEF | 18 | 22 | 35 | 41 |
| ALL | 9 | 13 | 25 | 37 |
| Average | 28.5 | 36.77 | 43.55 | 47.44 |

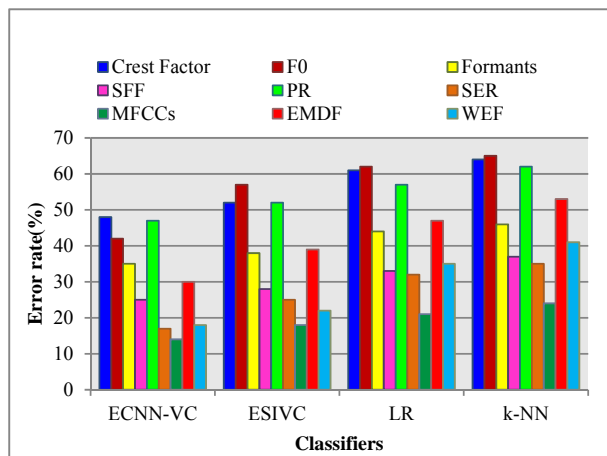


Figure 3. Error results vs. four classifiers without feature selection

The error results of the four different classifiers is illustrated in the fig 3, it concludes that the proposed ECNN-VC approach produces lesser error results of 9% which is

4%, 16% and 28 % lesser when compared to other classifiers with features sets. It concludes that the proposed work performs better than the other classifiers shown in table II.

E. After feature selection results

In this study, we use EFS algorithm for the multi-dimensional feature selection. Particularly, for all feature sets, the average performance of the proposed ECNN-VC approach considerably increases from 80% to 91%. For F0 (86), PR (80), and Crest Factor (85), improved after the completion of the EFS discussed in table 3. In the experiments results the proposed ECNN-VC approach achieve the UAR of 94% with the best combination of features.

Table 3. UAR ([%]) results obtained with nine features and four classifiers after EFS

| Features | ECNN-VC | ESIVC | LR | k-NN |
|---------------------|---------|-------|-------|-------|
| Crest Factor | 85 | 78 | 60 | 53 |
| F0 | 86 | 83 | 56 | 49 |
| Formants | 82 | 77 | 63 | 60 |
| SFF | 89 | 82 | 75 | 71 |
| PR | 80 | 74 | 63 | 58 |
| SER | 88 | 83 | 78 | 72 |
| MFCCs | 91 | 84 | 80 | 77 |
| EMDF | 82 | 75 | 68 | 65 |
| WEF | 90 | 85 | 75 | 70 |
| ALL | 94 | 91 | 85 | 82 |
| Average | 85.33 | 80.11 | 68.66 | 63.88 |

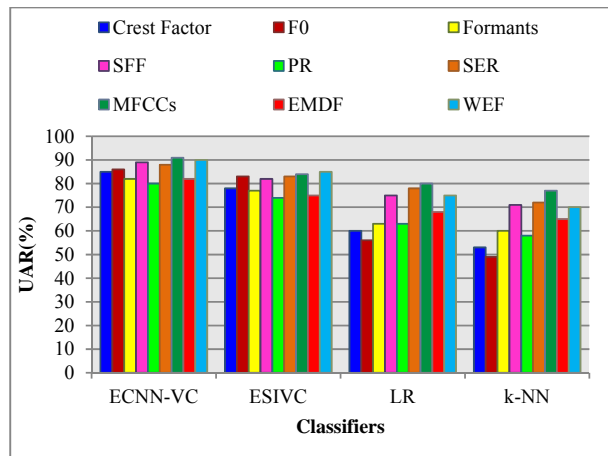


Figure 4. UAR results vs. four classifiers with EFS

Particularly, for all feature sets, the average performance of the proposed ECNN-VC approach produces 85.33% by EFS is illustrated in the fig 4. The average accuracy results of the proposed ESIVC approach is 85.33% which is 5.22%, 16.67% and 21.45% higher when compared to ESIVC, LR and k-NN classifiers respectively.

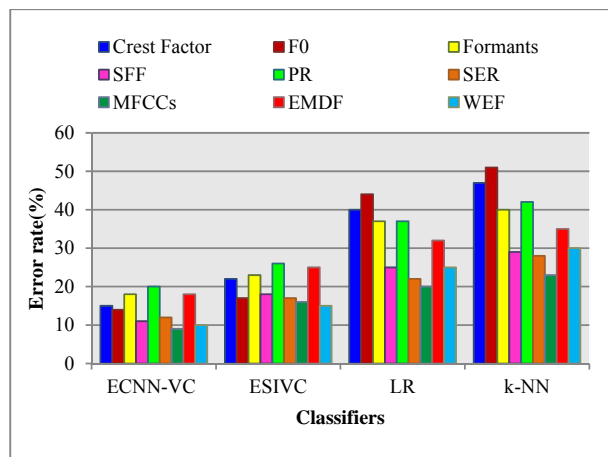


Figure 5. Error rate results vs. four classifiers with EFS

The error results of the four different classifiers with EFS is illustrated in the figure 5, it concludes that the proposed ECNN-VC approach produces lesser error results of 6% which is 3%, 9% and 12 % lesser when compared to other ESIVC, LR, and kNN classifiers respectively with all selected features sets. It concludes that the proposed ECNN-VC approach performs better than the other classifiers are discussed in table 4.

Table 4. Error rate ([%]) results obtained with nine features and four classifiers after EFS

| Features | ECNN-VC | ESIVC | LR | k-NN |
|--------------|---------|-------|----|------|
| Crest Factor | 15 | 22 | 40 | 47 |
| F0 | 14 | 17 | 44 | 51 |

| Formants | 18 | 23 | 37 | 40 |
|----------|-------|-------|-------|-------|
| SFF | 11 | 18 | 25 | 29 |
| PR | 20 | 26 | 37 | 42 |
| SER | 12 | 17 | 22 | 28 |
| MFCCs | 9 | 16 | 20 | 23 |
| EMDF | 18 | 25 | 32 | 35 |
| WEF | 10 | 15 | 25 | 30 |
| ALL | 6 | 9 | 15 | 18 |
| Average | 14.11 | 19.88 | 31.33 | 36.11 |

5. CONCLUSION AND FUTURE WORK

This paper introduces and experiment an Ensemble Feature Selection (EFS) and Ensemble Convolutional Neural Network with VOTE Classification (ECNN-VC) approach with high efficiency. In this work EFS approach, we combine the results of many FS algorithms, and introduce newly selected features within advanced ECNN-VC approach. The proposed ECNN-SVC approach showed the better detection results for snore sounds of 40 male patients has been recorded by using DISE, and classified by using Ear, Nose & Throat (ENT) experts. Regardless of a comparatively small data set, we are able to provide a higher detection performance with chosen feature sets self-regulating of individual subjects. The results demonstrated that the proposed EFS analysis and ECNN-VC approach provides promising results to assist recognizing the anatomical system of snore sound creation in particular subjects. Though the acoustics of snoring, as a diagnostic system is from a growing stage, there is an urgent need of new system to a accurate, huge data set, and a single snore occurrence test with an effectiveness measure with the purpose of returns the unbalanced features of snores to diagnose OSA.

REFERENCES

- [1] Peppard, P.E., Young, T., Barnet, J.H., Palta, M., Hagen, E.W. and Hla, K.M., 2013. Increased prevalence of sleep-disordered breathing in adults. *American journal of epidemiology*, 177(9), pp.1006-1014.
- [2] Rakel, R.E., 2009. Clinical and societal consequences of obstructive sleep apnea and excessive daytime sleepiness. *Postgraduate medicine*, 121(1), pp.86-95.
- [3] Hamilton, G.S., Meredith, I.T., Walker, A.M. and Solin, P., 2009. Obstructive sleep apnea leads to transient uncoupling of coronary blood flow and myocardial work in humans. *Sleep*, 32(2), pp.263-270.

- [4] Calvin AD, Somers VK. Obstructive sleep apnea and risk of stroke: time for a trial. *Nat Clin Pract Cardiovasc Med* 2009;6:90–1.
- [5] Pizza, F., Contardi, S., Mondini, S., Trentin, L. and Cirignotta, F., 2009. Daytime sleepiness and driving performance in patients with obstructive sleep apnea: comparison of the MSLT, the MWT, and a simulated driving task. *Sleep*, 32(3), pp.382-391.
- [6] Flemons, W.W. and Tsai, W., 1997. Quality of life consequences of sleep-disordered breathing. *Journal of allergy and clinical immunology*, 99(2), pp.S750-S756.
- [7] Kushida, C.A., Littner, M.R., Morgenthaler, T., Alessi, C.A., Bailey, D., Coleman Jr, J., Friedman, L., Hirshkowitz, M., Kapen, S., Kramer, M. and Lee-Chiong, T., 2005. Practice parameters for the indications for polysomnography and related procedures: an update for 2005. *Sleep*, 28(4), pp.499-523.
- [8] Ghaemmghami, H., Abeyratne, U.R. and Hukins, C., 2009, Normal probability testing of snore signals for diagnosis of obstructive sleep apnea. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009. EMBC 2009. pp. 5551-5554.
- [9] Finkel, K.J., Searleman, A.C., Tymkew, H., Tanaka, C.Y., Saager, L., Safer-Zadeh, E., Bottros, M., Selvidge, J.A., Jacobsohn, E., Pulley, D. and Duntley, S., 2009. Prevalence of undiagnosed obstructive sleep apnea among adult surgical patients in an academic medical center. *Sleep medicine*, 10(7), pp.753-758.
- [10] D. Pevernagie, R. M. Aarts, and M. De Meyer, "The acoustics of snoring," *Sleep Medicine Reviews*, vol. 14, no. 2, pp. 131–144, 2010.
- [11] Roebuck, V. Monasterio, E. Geder, M. Osipov, J. Behar, A. Malhotra, T. Penzel, and G. Clifford, "A review of signals used in sleep analysis," *Physiological Measurement*, vol. 35, no. 1, pp. R1–R57, 2014.
- [12] Bliwise, D. L., Nekich, J. C. & Dement, W. C. Relative validity of self-reported snoring as a symptom of sleep apnea in a sleep clinic population. *Chest* 99, 600–608 (1991).
- [13] Nakano, H., Furukawa, T. and Nishima, S., 2008. Relationship between snoring sound intensity and sleepiness in patients with obstructive sleep apnea. *Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine*, 4(6), pp. 551–556.
- [14] Ng, A.K., San Koh, T., Abeyratne, U.R. and Puvanendran, K., 2009. Investigation of obstructive sleep apnea using nonlinear mode interactions in nonstationary snore signals. *Annals of biomedical engineering*, 37(9), pp.1796-1806.
- [15] Ng, A.K., San Koh, T., Baey, E., Lee, T.H., Abeyratne, U.R. and Puvanendran, K., 2008. Could formant frequencies of snore signals be an alternative means for the diagnosis of obstructive sleep apnea?. *Sleep medicine*, 9(8), pp.894-898.
- [16] Alencar, A.M., da Silva, D.G.V., Oliveira, C.B., Vieira, A.P., Moriya, H.T. and Lorenzi-Filho, G., 2013. Dynamics of snoring sounds and its connection with obstructive sleep apnea. *Physica A: Statistical Mechanics and its Applications*, 392(1), pp.271-277.
- [17] Mesquita, J., Solà-Soler, J., Fiz, J.A., Morera, J. and Jané, R., 2012. All night analysis of time interval between snores in subjects with sleep apnea hypopnea syndrome. *Medical & biological engineering & computing*, 50(4), pp.373-381.
- [18] Ben-Israel, N., Tarasiuk, A. and Zigel, Y., 2012. Obstructive apnea hypopnea index estimation by analysis of nocturnal snoring signals in adults. *Sleep*, 35(9), pp.1299-1305.
- [19] Fiz, J.A., Jane, R., Solà-Soler, J., Abad, J., García, M. and Morera, J., 2010. Continuous analysis and monitoring of snores and their relationship to the apnea-hypopnea index. *The Laryngoscope*, 120(4), pp.854-862.
- [20] Collop, N.A., Tracy, S.L., Kapur, V., Mehra, R., Kuhlmann, D., Fleishman, S.A. and Ojile, J.M., 2011. Obstructive sleep apnea devices for out-of-center (OOC) testing: technology evaluation. *Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine*, 7(5), pp.531–48.
- [21] C. Janott, W. Pirsig, and C. Heiser, "Akustische analyse von schnarchger "auschen," *Somnologie-Schlafforschung und Schlafmedizin*, vol. 18, no. 2, pp. 87–95, 2014.
- [22] Jin, H., Lee, L.A., Song, L., Li, Y., Peng, J., Zhong, N., Li, H.Y. and Zhang, X., 2015. Acoustic analysis of snoring in the diagnosis of obstructive sleep apnea syndrome: a call for more rigorous studies. *Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine*, 11(7), p.765.
- [23] Xu, H., Zheng, X., Qian, Y., Guan, J., Yi, H., Zou, J., Wang, Y., Meng, L., Zhao, A., Yin, S. and Jia, W., 2016. Metabolomics profiling for obstructive sleep apnea and simple snorers. *Scientific reports*, 6, p.30958.
- [24] A. S. Karunajeewa, U. R. Abeyratne, and C. Hukins, "Multi-feature snore sound analysis in obstructive sleep apnea-hypopnea syndrome," *Physiological Measurement*, vol. 32, no. 1, pp. 83–97, 2011.
- [25] Saha, S., Bradley, T.D., Taheri, M., Moussavi, Z. and Yadollahi, A., 2016. A subject-specific acoustic model of the upper airway for snoring sounds generation. *Scientific reports*, 6, p.25730.

- [26] Abeyratne, U.R., De Silva, S., Hukins, C. and Duce, B., 2013. Obstructive sleep apnea screening by integrating snore feature classes. *Physiological measurement*, 34(2), pp.99-121.
- [27] Praydas T, Wongkittisuksa B, and Tanthanuch S, "Obstructive Sleep Apnea Severity Multiclass Classification Using Analysis of Snoring Sounds", *Proceedings of the 2nd World Congress on Electrical Engineering and Computer Systems and Science (EECSS'16)*, pp.1-5.
- [28] Qian, K., Janott, C., Pandit, V., Zhang, Z., Heiser, C., Hohenhorst, W., Herzog, M., Hemmert, W. and Schuller, B., 2017. Classification of the Excitation Location of Snore Sounds in the Upper Airway by Acoustic Multifeature Analysis. *IEEE Transactions on Biomedical Engineering*, 64(8), pp.1731-1741.
- [29] Dedhia, R.C. and Weaver, E.M., 2015. Association between drug-induced sleep endoscopy and measures of sleep apnea burden. *Otolaryngology–Head and Neck Surgery*, 153(5), pp.875-880.
- [30] M. S. Aldrich, *Sleep Medicine*. Transaction Publishers, 1999.
- [31] E. J. Kezirian, W. Hohenhorst, and N. de Vries, "Drug-induced sleep endoscopy: the vote classification," *European Archives of Oto-Rhino- Laryngology*, vol. 268, no. 8, pp. 1233–1236, 2011.
- [32] K. Qian, Y. Fang, Z. Xu, and H. Xu, "Comparison of two acoustic features for classification of different snore signals," *Chinese Journal of Electron Devices*, vol. 36, no. 4, pp. 455–459, 2013.
- [33] J. R. Deller Jr, J. G. Proakis, and J. H. Hansen, *Discrete Time Processing of Speech Signals*. Prentice Hall PTR, 1993.
- [34] K. Qian, Z. Xu, H. Xu, and B. P. Ng, "Automatic detection of inspiration related snoring signals from original audio recording," in *Proc. of the IEEE ChinaSIP*. Xi'an, China: IEEE, 2014, pp. 95–99.
- [35] A. K. Ng, T. San Koh, K. Puvanendran, and U. R. Abeyratne, "Snore signal enhancement and activity detection via translation-invariant wavelet transform," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 10, pp. 2332–2342, 2008.
- [36] R. N. Khushaba, S. Kodagoda, S. Lal, and G. Dissanayake, "Driver drowsiness classification using fuzzy wavelet-packet-based feature extraction algorithm," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 121–131, 2011.
- [37] Saeys, Y., Abeel, T. and Van de Peer, Y., 2008, Robust feature selection using ensemble feature selection techniques. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* , pp. 313-325.
- [38] Saeys, Y., Inza, I. and Larrañaga, P., 2007. A review of feature selection techniques in bioinformatics. *bioinformatics*, 23(19), pp.2507-2517.
- [39] Colla, V., Matarese, N., Reyneri M.L. A method to point out anomalous input-output patterns in a database for training neurofuzzy system with a supervised learning rule. *9th International Conference on Intelligent Systems Design and Applications ISDA 2009*, pp. 1307-1311.
- [40] Golub T, Slonim D, Tamayo P, Huard C, Gaasenbeek M, Mesirov J, Coller H, Loh M, Downing J, Caligiuri M, Lander CBE (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286:531–537.
- [41] Maldorad S, Weber R (2009) A wrapper method for feature selection using support vector. machines. *Information Sciences* 179:2208–2217.
- [42] Gu Q, Li Z, Han J (2011) Generalized fisher score for feature selection. *Proc Conf Uncertainty in artificial Intelligence* 1:266–273
- [43] Rice JA (2006) *Mathematical Statistics and Data Analysis*. Third Edition.
- [44] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).
- [45] Izadyyazdanabadi, M., Belykh, E., Mooney, M., Martirosyan, N., Eschbacher, J., Nakaji, P., Preul, M.C. and Yang, Y., 2017. Convolutional Neural Networks: Ensemble Modeling, Fine-Tuning and Unsupervised Semantic Localization. *arXiv preprint arXiv:1709.03028*.
- [46] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S. and Darrell, T., 2014, November. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 675-678). ACM.
- [47] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: *Advances in neural information processing systems*, 2012, pp. 1097–1105

Leaf Recognition System for Plant Identification and Classification Based On Leaf Features Fusion & Zernike Moment

Jignesh B. Jethva¹, PhD Scholar,
Prof. Dr. C. R. Singla²

^{1,2} Electronics and Communication Department,

Madhav University, Abu Road, Pindwara, Dist. Sirohi, Bharja, Rajasthan, INDIA – 307026

Jethvajignesh1982@gmail.com¹
crslibra1010@gmail.com²

Abstract- Imagine someone hiking in the Swiss mountains, where he finds a weird leaf or flower. This person has always been bad in biology but would like to know more about that plant. What's its name? What are their main features? Is it rare? Is it protected? etc. By simply taking a picture of the leaf with a Digital Camera, he or she could feed it to the database in his computer and then get all the information regarding the leaf image through an automatic leaf recognition application.

Even today, identification and classification of unknown plant species are performed manually by expert personnel who are very few in number. The important aspect is to develop a system which classifies the plants. This paper presents a new recognition approach based on Leaf Features Fusion and Random Forests (RF) Classification algorithms for classifying the different types of plants. The proposed approach consists of three phases that are pre-processing, feature extraction, and classification phases. Since most types of plants have unique leaves. Leaves are different from each other by characteristics such as the shape, color, texture and the margin.

This is an intelligent system which has the ability to identify tree species from photographs of their leaves and it provides accurate results in less time.

Keywords: Random Forest, Zernike Moment, Gabor Filters, GLCM

Introduction

Since late decades, computerized picture preparing, picture investigation & machine vision have been forcefully created, and they have turned into a vital piece of manmade brainpower and the interface amongst human & machine grounded hypothesis and connected innovation. These innovations have been connected generally in industry & drug, however seldom in domain identified with horticulture or normal environments.

“A standout amongst the most imperative assignments for researchers, field aides, and others are order of plants, since plants have a critical part in the characteristic hover of life. They are key to practically every other type of life, as they shape the biggest part of the living life forms that can change over the daylight into nourishment. What's more, as all oxygen noticeable all around that people and different creatures inhale is delivered by plants, thus without plants it is hard to consider presence of human life on earth. Characterizing plants helps at guaranteeing the security and survival of all normal life. The procedure of plant characterization can be

performed utilizing distinctive routes, for example, cell and atomic science and in addition utilizing the plants' takes off".

"Most sorts of plants have special leaves that are not quite the same as each other in light of various attributes, for example, shape, shading, surface, and the edge. The generous data conveyed by each can be utilized to recognize and characterize the inception or the kind of plant, so leaf acknowledgment/order is essential assignment at the procedure of plant characterization".

"Lately, there has been considerable work in the PC vision field, which handled the issue of plants arrangement utilizing leafs acknowledgment. One can without much of a stretch exchange the leaf picture to a PC and a PC can remove highlights consequently in picture handling systems. A few frameworks utilize depictions utilized by botanists. Be that as it may, it is difficult to concentrate and exchange those components to a PC naturally".

The main goal of this research is to create a Leaf recognition program based on specific characteristics extracted from photography. Hence this presents an approach where the plant can be classified based on their leaf features such as color, shape & texture and classification. The main purpose of this program is to use MATLAB resources.

Motivation

The human visual system has no problem interpreting the subtle variations in translucency and shading in this Figure 1. Photograph and correctly segmenting the object from its background.



Figure 1. Lotus flower seen as to the naked eye.

Let's imagine a person taking a field trip, and seeing a bush or a plant on the ground, he or she would like to know whether it's a weed or any other plant but have no idea about what kind of plant it could be. With a good digital camera and arecognition program, one could get some useful information. "Plants assume a vital part in our surroundings. Without plants there will be no presence of the world's nature. Be that as it may, as of late, many sorts of plants are at the danger of termination. To ensure plants and to list different sorts of greenery diversities, a plant database is a critical stride towards protection of earth's biosphere. There are a colossal number of plant species around the world. To handle such volumes of data, improvement of a snappy and effective characterization technique has turned into a region of dynamic research. Notwithstanding the preservation angle, acknowledgment of plants is additionally important to use their restorative properties and utilizing them as wellsprings of option vitality sources like bio-fuel. There are a few approaches to perceive a plant, similar to bloom, root, and leaf, organic product and so on".

Existing Works

This section describes the previous work which had been done for Leaf Identification.

Pallavi P et al., (and other) [1] developed "a new structure for perceiving and distinguishing plants is been proposed. Shape, vein, shading and surface components have been utilized to recognize the leaf and neural

system approach is utilized to arrange them. In this, GLCM gives better surface approximations and thus makes arrangement simpler”.

OluleyeBabatunde et al. [2] demonstrate the different systems close by their portrayals. It depicts how future analysts in this field may advance the learning area.

Stephen Gang Wu et al. [3] utilize “Probabilistic Neural Network (PNN) with picture and information handling systems to execute broadlyuseful robotized leaf acknowledgment for plant grouping. 12 leaf elements are separated and orthogonalized into 5 chief factors which comprise the info vector of the PNN. The PNN is prepared by 1800 leaves to group 32 sorts of plants with precision more noteworthy than 90%”.

AnandHanda et al. [4] finish up with the continuous work in the present zone and the other existing issues in the region. “The programmed advanced plant grouping should be possible by removing different elements from its leaves and still there exist potential outcomes to enhance plant species distinguishing proof through the planning of another computerized programmed plant recognizable proof and acknowledgment framework”.

M. M. Amlekar et al. [5] different administrators are examined for the leaf extraction from pictures by utilizing the picture handling strategies.

A Gopal et al. [6] prepare product with as well (10 number of every plant species) leaves and tried with 50 (tried with various plant species) clears out. The efficiency of system is to be 92%.

EsraaElhariri et al. [7] introduce “a grouping approach in view of RF and LDA calculations for characterizing the distinctive sorts of plants. Leaves are not quite the same as each other by qualities, for example, the shape, shading, surface and the edge. LDA accomplished characterization exactness of (92.65 %) against the RF that accomplished precision of (88.82 %) with mix of shape, first request surface, Gray Level Co-event Matrix (GLCM), HSV shading minutes, and vein highlights”.

AnantBhardwaj et al. [8] displayed “different successful calculations utilized for plant order utilizing leaf pictures and audit the principlecomputational, morphological and picture preparing techniques that have been utilized as a part of late years”.

BoranSekeroglua et al. [9] presented “intelligent recognition system to recognize and identify 27 different types of leaves using back propagation neural network and results show that the developed system is superior to recent researches with the recognition rate of 97.2%”.

Rongxiang Hu, Wei Jia et al. [10] connected “the proposed strategy to the undertaking of plan leaf acknowledgment with trials on two datasets: the Swedish Leaf dataset and the ICL Leaf dataset”.

TrishenMunisami et al. [11] Developed “a mobile application to allow a user to take pictures of leaves and upload them on server. The server runs pre-processing and feature extraction techniques on the image before a pattern matcher compares information from this image with the ones in database in order to get potential matches”.

AjinkyaGawade et al. [12] are attempting to acquire atomization this procedure to such an extent that with no past learning of the leaf species to layman simply utilizing its picture.

Sachin D et al. [13] present “a computer based automatic plant identification system. Out of all available organs of plant, leaf is selected to obtain the features of plant. Five geometrical parameters are calculated using digital image processing techniques. On the basis of these geometrical parameters six basic morphological features are extracted. Vein feature as a derived feature is extracted based on leaf structure”.

Miss. NeedaSamreen I et al. [14] discusses “the leaf recognition which enables the user to recognize the type of leaf using a approach that depends on neural network. Scanned images are being introduced into the computer initially, image enhancement and reduction of noise modifies their quality, further followed by feature extraction”.

XiaoWei Shao et al. [15] another sort of detecting gadget, the Kinect profundity sensor which measures the genuine separation to objects straightforwardly and can catch high-determination profundity pictures, is abused for the programmed acknowledgment and extraction of takes off.

Arunpriya C et al. [16] comprises of three stages, for example, preprocessing, include extraction and characterization to prepare the stacked picture. The tea leaf pictures can be distinguished precisely in the preprocessing stage by fluffy denoising utilizing Dual Tree Discrete Wavelet Transform (DT-DWT). In the component extraction stage, Digital Morphological Features (DMFs) are inferred to enhance the grouping precision.

KshitijFulsoundar et al. [17] portray the improvement of an Android application that gives clients the capacity to recognize plant species in light of photos of the plant's leaves brought with a cell phone. The Core of this system is a calculation that secures morphological components of the leaves, registers very much archived measurements.

JyotismitaChakia et al. [18] show “a new strategy of portraying and perceiving plant leaves utilizing a blend of surface and shape highlights. Surface of the leaf is demonstrated utilizing Gabor channel and Gray Level Co-event Matrix (GLCM)”.

ShyamVijayraoPundkar et al. [19] demonstrate that picture handling is driving area in recognizable proof of restorative plant.

Deore Nikita R et al. [20] use “mobile phones for real time monitoring of plant disease for proper diagnosis and treatment. A central server is placed at the pathological laboratory for sharing of the data collected by the mobile phones”.

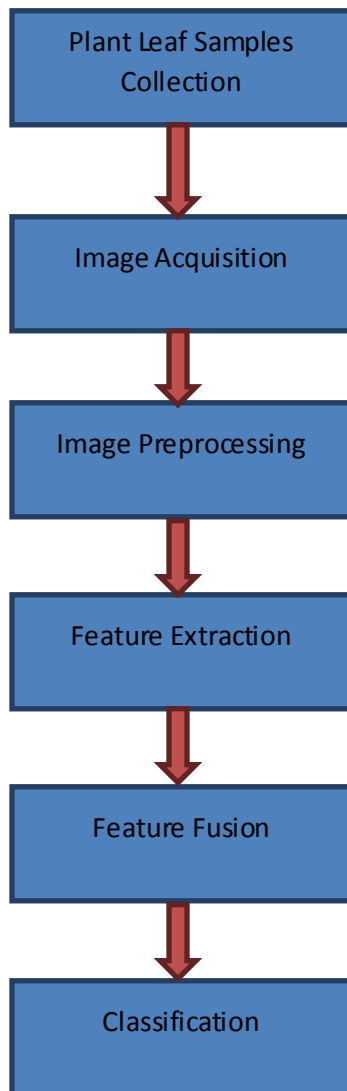
Proposed Work

In Our Proposed Approach, First step is Leaf Image Acquisition. In this step digital leaf image is captured. Then apply pre-processing step on these leaf image. Prior to the operations, a portion of the leaf pictures are turned physically to help the program to mastermind leaf zenith bearing to the correct side. A short time later, programmed pre-handling procedures are connected to the greater part of the leaf pictures. Pre-processing steps involves converting RGB to Grayscale Image, then apply Median Filtering on it, then converting into binary and apply segmentation on it. After pre-processing, the important and essential task is to measure the properties of an object which is called Feature Extraction because objects have to be detected based on these computed

properties. In Feature extraction, I will extract Features such as Color Features, Texture Features, Shape Features, and Vein Features and also apply Zernike Moments of Leaf Image. After Feature Extraction, next step is Feature Fusion to combine more than one Feature to get more accuracy for classification.

Once the features have been fused, then these features vectors are to be used to classify and identify plant using RF (Random Forest) classifier to classify plants. A brief explanation on the proposed system is given in the Figure 2.

Training Phase



Testing Phase

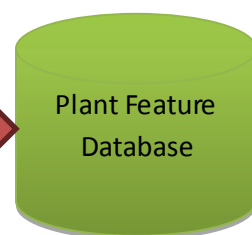
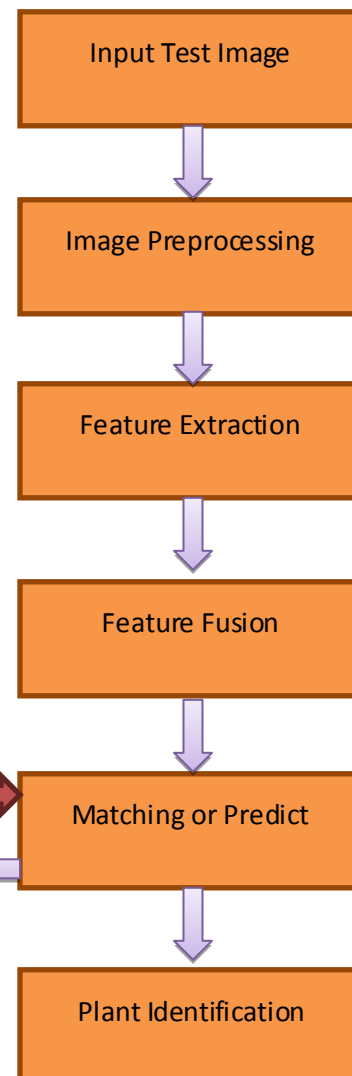


Figure 2. Proposed Work

A. Algorithm for Proposed System

Steps:

1. Prepare Training Dataset
 - 1.1. Collect Plant Leaf Samples.
 - 1.2. Acquisition of plant leaf images
 - 1.3. Apply Preprocessing on each plant leaf image includes Gray conversion, median filtering and then binarization and segmentation.
 - 1.4. Extract Features of plant leaf such as shape, color, vein, texture etc. and apply Zernike moment.
 - 1.5. Fuse the features based on combination.
 - 1.6. Prepare features vector.
2. Read the testing plant leaf image
3. Apply Pre-processing on test image including same steps in step 1.3
4. Extract Features specified in Step 1.4 and Fuse them based on combination.
5. Train the training dataset and predicate testing image by using Random Forest Classifier.
6. Finally, identify the plant leaf.
7. Stop.

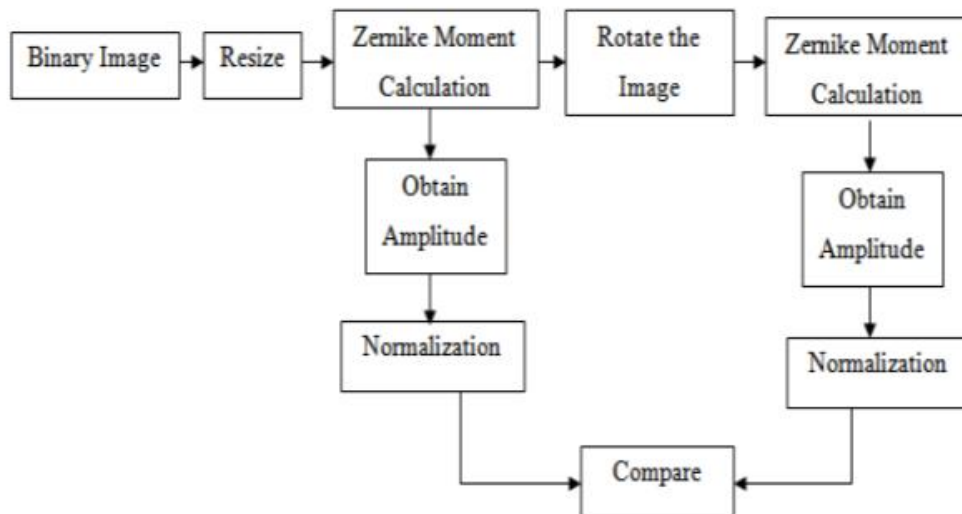
B. Features Extraction In Proposed Work

| Features | Sub Features | Description |
|-----------------------|---------------------------------|---|
| Shape Features | Eccentricity | “It is defined as the ratio of the distance between the foci of the ellipse and its major axis length. It is used to differentiate rounded and long leaves” |
| | Solidity | “It is the ratio between object's area and area of the object's convex hull. It may be considered as a certain measure of convexity“ $\text{Solidity} = \frac{A(I)}{A(H(I))}$ Where A(I) is the object area and A(H(I)) is the area of object's convex hull. |
| | Aspect Ratio (AR) | “It is the ratio between the maximum length D_{MAX} and the minimum length D_{MIN} of the minimum bounding rectangle”. $AR = \frac{D_{MAX}}{D_{MIN}}$ |
| | Width Ratio (WR) | “It is the ratio of width at half of major axis to maximum width”. |
| | Perimeter | “It is scalar that specifies the distance around boundary of the region” |
| | Area | “It is scalar that specifies the actual number of pixels in the region” |
| | Roundness or Circularity | “It is the ratio of $4*PI*Area$ of the leaf to the square of perimeter” |

| | | |
|-------------------------|---|--|
| | EquivDiameter | “It is scalar that specifies the diameter of a circle with same area as the region computed as $\sqrt{4 \cdot \text{Area} / \pi}$ ” |
| | Centroid | “It is 1 by Q vector that specifies the center of mass of region. First element of centroid is horizontal coordinate of center of mass and second element is the vertical coordinate” |
| | Convex Area | “It is scalar that specifies the number of pixels in convex Image” |
| | Convex Hull | “It is p by 2 matrixes that specify smallest convex polygon that contain the region” |
| Color Features | Mean | $\bar{X}_i = \frac{\sum_{j=1}^{MN} X_{i,j}}{M \cdot N}$ |
| | Standard deviation | $\sigma_i = \sqrt{\frac{1}{M \cdot N} \sum_{j=1}^{MN} (X_{i,j} - \bar{X}_i)^2}$ |
| | Skewness | $S_i = \sqrt[3]{\frac{1}{M \cdot N} \sum_{j=1}^{MN} (X_{i,j} - \bar{X}_i)^3}$ |
| | Kurtosis | $K_i = \sqrt[4]{\frac{1}{M \cdot N} \sum_{j=1}^{MN} (X_{i,j} - \bar{X}_i)^4}$ |
| | Where $X_{i,j}$ is the value of image pixel j of color channel i. \bar{X}_i is the mean for each channel i. σ_i is the standard deviation, S_i is skewness and K_i is kurtosis for each channel | |
| Vein Features | “Vein features are features derived from vein of the leaf. There are four kinds of vein features, defined as follows: $V1=A1/A$, $V2=A2/A$, $V3=A3/A$, $V4=A4/A$ Where $A1$, $A2$, $A3$ and $A4$ are pixel number that constructs the vein and A is area of the leaf”. | |
| Texture Features | GrayLevel Co-occurrence Matrix (GLCM) | $\text{Angular Second Moment} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P_{ij}^2$ $\text{Contrast} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i-j)^2 P_{ij}$ $\text{Correlation} = \frac{1}{\sigma_x \sigma_y} \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} [ij P_{ij} - \mu_x \mu_y]$ $\text{Entropy} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P_{ij} \log P_{ij}$ $\text{variance} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i - \mu)^2 P_{ij}$ $\text{Homogeneity} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{1}{(i-j)^2} P_{ij}$ |

| | | |
|--|---------------------|---|
| | | $\text{Sum of Entropy} = - \sum_{i=2}^{2G-2} P_{x+y}(i) \log P_{x+y}(i)$ $\text{Cluster Shade} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i+j-\pi_x-\pi_y)^3 P_{ij}$ $\text{Prominence} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i+j-\pi_x-\pi_y)^4 P_{ij}$ <p>Where μ_x, μ_y, σ_x and σ_y mean and standard deviation of corresponding distribution and G is number of Gray levels</p> |
| | Gabor Filter | <p>A complex Gabor filter is defined as the product of a Gaussian kernel and a complex sinusoid. A 2D Gaussian curve g with a spread of σ in both x and y directions”, is represented as below:</p> $g(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{x^2 + y^2}{2\sigma^2}\right]$ $(x, y, u, \theta, \varphi) = \exp\{j2\pi(x.u\cos\theta + y.u\sin\theta) + \varphi\}$ |

C. Zernike Moment



Performance Parameters

The performance of the proposed system is tested with Random Forest classifiers by using the feature set extracted in the dataset. Confusion matrix, sensitivity, accuracy, kappa statistics, RMSE and AUROC metrics are measured

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$RMSE = \sqrt{\frac{\sum (y' - y)^2}{n}}$$

Here TP , TN , FP and FN denote the number of Leaf Images classified as true positive, true negative, false positive, and false negative, respectively. In the root mean squared error ($RMSE$), y and y' depict actual and predicted values. n is the number of Leaf images.

TP (True Positives) refers to the positive images that were correctly labelled by the classifier.

TN (True Negatives) refers to the negative images that were correctly labelled by the classifier.

FP (False Positives) refers to the negative images that were incorrectly labelled as positive by the classifier.

FN (True Positives) refers to the positive images that were incorrectly labelled as negative by the classifier.

Conclusion

We conclude that incorporating Zernike moments for feature descriptors is a feasible alternative for classifying structurally complex images. They offer exceptional invariance features and reveal enhanced performance than other moment based solutions. Gabor and GLCM give better texture approximations and hence make classification easier. Random Forest Classifier gives better accuracy than any other classifier. We have used features fusion with Zernike moments to recognize plant leaf with accuracy more than 98%.

References

1. "Leaf Recognition Based on Feature Extraction and Zernike Moments", Pallavi P, V.S Veena Devi, International Journal of Innovative Research in Computer and Communication Engineering, May 2014.
2. "A survey of computer-based vision systems for automatic identification of plant species", Oluleye Babatunde, Leisa Armstrong, Dean Diepeveen, Jinsong Leng, Journal of Agricultural Informatics, 2015.
3. "A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network", Stephen Gang Wu, Forrest Sheng Bao, Eric You Xu, Yu-Xuan Wang, Yi-Fan Chang and Qiao-Liang Xiang, International Symposium on Signal Processing and Information Technology, 2007
4. "A Review and a Comparative Study of Various Plant Recognition and Classification Techniques using Leaf Images", Anand Handa And Rashmi Agarwal, International Journal of Computer Applications, August 2015
5. "Leaf Shape Extraction For Plant Classification", M. M. Amlekar, A. T. Gaikwad, R. R. Manza, P. L. Yannawar, 2015 International Conference on Pervasive Computing (ICPC), 2015
6. "Classification of selected medicinal plants leaf using image processing", Gopal, A, S. Prudhveeswar Reddy, and V. Gayatri. 2012 International Conference on Machine Vision and Image Processing (MVIP), 2012.
7. "Plant classification system based on leaf features", Elhariri, Esraa, Nashwa El-Bendary, and Aboul Ella Hassanien, 2014 9th International Conference on Computer Engineering & Systems (ICCES), 2014
8. "A review on plant recognition and classification techniques using leaf images", Anant Bhardwaj, Manpreet Kaur, International Journal of Engineering Trends and Technology - Volume 4 Issue 2- 2013
9. "Leaves recognition system using a neural network", Boran Sekeroglu, Yucel Lnan, 12th International Conference on Application of Fuzzy Systems and Soft Computing, ICAFS, 2016

10. "Multiscale Distance Matrix for Fast Plant Leaf Recognition", Hu, Rongxiang, Wei Jia, Haibin Ling, and De- Shuang Huang. IEEE Transactions on Image Processing, 2012.
11. "Plant leaf recognition using shape features and colour histogram with k-nearest neighbour classifiers", TrishenMunisami, MaheshRamsum, SomveerKishnah, SameerchandPudaruth, Second International Symposium on Computer Vision and the Internet, Elsevier, 2015
12. "Automated Tool for Plant Leaf Identification using Morphological Features", AjinkyaGawade, AnandSartape, AamodChemburkar, Prasad Somawanshi, ReenaPagare, JayshreeGhorpade, International Journal of Emerging Research in Management &Technology, May 2015
13. "Plant Identification Using Leaf Images.", Sachin D. Chothe, V.R.Ratnaparkhe, International Journal of Innovative Research in Science, Engineering and Technology, May 2015
14. "A Review on Off-line Leaf Recognition Using Neural Network", Miss. NeedaSamreenI.Khan, Prof. Rajesh B.Pandhare, International Journal of Computer Science and Mobile Computing, Jan 2015
15. "Leaf recognition and segmentation by using depth image", Shao, Xiaowei, Yun Shi, Wenbing Wu, Peng Yang, Zhongxin Chen, and RyosukeShibasaki, 2014 The Third International Conference on Agro-Geoinformatics, 2014
16. "An Effective Tea Leaf Recognition Algorithm for Plant Classification Using Radial Basis Function Machine", ArunpriyaC., Antony SelvadossThanamani, International Journal of Modern Engineering Research (IJMER), Mar. 2014
17. "Detection and classification of plant leaf diseases", KshitijFulsoundar, TusharKadlag, SanmanBhadale, Pratik Bharvirkar, Prof S.P.Godse, International Journal of Engineering Research and General Science Volume 2, Issue 6, October-November, 2014
18. "Plant Leaf Recognition using Texture and Shape features with Neural Classifiers", JyotismitaChakia, RanjanParekha, and Samar Bhattacharyaa, Pattern Recognition Letters, 2015
19. "Study of Various Techniques for Medicinal Plant Identification", ShyamVijayraoPundkar, Prof. M.M. Waghmare, International Journal on Recent and Innovation Trends in Computing and Communication, November 2014
20. "Plant Leaf Disease Identificationand Prevention by Mobile Vision System", Deore Nikita R., KhadeChhaya A., KotwalPooja R.&PatilKalyani J, Imperial Journal of Interdisciplinary Research (IJIR) , 2016

A Hybrid Approach for Classification of MRI Brain Tumors Using Genetic Algorithm, K-Nearest Neighbor and Probabilistic Neural Network

Raghad Majeed Azawi

Department of Computer Science
College of Science, University of Diyala
Diyala, Iraq
raghadma2016@yahoo.com

Dhahir Abdulhade Abdulah

Department of Computer Sciences
College of Science, University of Diyala
Diyala, Iraq
Dhahir@yahoo.com

Jamal Mustafa Abbas

Department of Computer Science
College of Science, University of Diyala
Diyala, Iraq
altuwaijari@yahoo.com.uk

Ibrahim Tareq Ibrahim

College of Medicine, University of
Diyala, Diyala, Iraq
Ibr_33@yahoo.com

Abstract—Detection and classification of brain tumor are very important because it provides anatomical information of normal and abnormal tissues which helps in early treatment planning and patient's case follow-up. There is a number of techniques for medical image classification. We used PNN (Probabilistic Neural Network Algorithm) for image classification technique based on Genetic Algorithm (GA) and K-Nearest Neighbor (K-NN) classifier for feature selection is proposed in this paper. The searching capabilities of genetic algorithms are explored for appropriate selection of features from input data and to obtain an optimal classification. The method is implemented to classify and label brain MRI images into seven tumor types. A number of texture features (Gray Level Co-occurrence Matrix (GLCM)) can be extracted from an image, so choosing the best features to avoid poor generalization and over specialization is of paramount importance then the classification of the image and compare results based on the PNN algorithm.

Keywords - Brain tumors, MRI, Gray Level Co-occurrence Matrix (GLCM), Classification accuracy, Genetic Algorithm (GA), K-Nearest Neighbor (K-NN) and Probabilistic Neural Network Algorithm (PNN).

I. INTRODUCTION

The human body is made of many cells. Each cell has a specific job. The cells grow within the body and are divided to reproduce new cells. These divisions have certain functions in the body. But when each cell loses the ability to control its growth, these divisions are done without any limitations, and tumor consists. The brain is the central part of the human body responsible for coordinating and observing all other body organs, so if a tumor is present in any part of the brain then the activities controlled by this part of the nervous system are also affected. There are two types of brain tumors

malignant tumor and benign tumor [1]. Many imaging techniques can be used to diagnose and detect brain tumors early. Compared to all other imaging techniques, MRI is actively involved in the application of brain tumor identification and detection. It does not use ionizing radiation (X-rays) [2].

II. LITERATURE REVIEW

- **N.D.Pergad and Kshitija V.Shingare in 2015 [4]** designed a system for brain tumor extraction. This proposed system consists of preprocessing method for removing noise and Gray Level Co-occurrence Matrix (GLCM) for feature extraction step. Probabilistic Neural Network (PNN) is used for classification step of the image into normal and abnormal tumors. The last step is segmentation technique. The accuracy of this proposed system is 88.2%.
- **Naveena H. S. et al., in 2015[5]** exploited the capability of ANN algorithm for classification of MRI brain tumor images to either cancerous or non-cancerous. K-means clustering algorithm was used for segmentation stage. Then, gray level co-occurrence matrix (GLCM) was used for feature extraction stage of segmented image. Finally, Backpropagation neural network (BPNN) and Probabilistic Neural Network (PNN) is used for classification stage of brain tumors. The overall accuracy of the system is 79.02% in case of BPNN algorithm and 97.25% in case of PNN algorithm.
- **Ata'a A. and Dhia A. in 2016 [3]** this system is to detect and define tumor type in MRI brain images. The proposed system consists of multiple phases. The preprocessing stage the MRI image. Step two, transformations (features extraction algorithm based on using two level of 2-D discrete wavelet (DWT) and multiwavelet (DMWT) decomposition). Step three, the statistical measurements utilized to extract features from (GLCM). Step four, which deals with classification utilized (PNN) and the final Step, a proposed algorithm to segment, Superpixel Hexagonal Algorithm. The accuracy of testing in DWT is 91% and in case DMWT is 97%.
- **S.U Aswathy, and et.al, in 2017 [21]** designed a system for brain tumor segmentation using a genetic algorithm with SVM classifier. The proposed system is consisting of multiple phases. Step one is Pre-processing using the high pass, low pass and median filter for preprocessing. Step two, the segmentation by using a combination of expectation maximization (EM) algorithm and the level set method. Step three, feature extraction and selection using GA. Step four, classification MRI brain image to normal or abnormal by using SVM. The present work segments the tumor using Genetic Algorithm and classification of the tumor by using the SVM classifier.

III. THE PROPOSED SYSTEM

In the proposed system seven types of MRI image (normal and six types of tumors are considered, these are Lymphoma, Glioblastoma multiform, Cystic oligodendroglioma, Ependymoma, Meningioma and Anaplastic astrocytoma). The input data set consists of 140 (20 images for each type of the six tumors and 20 images for normal images) with 8 bit (pixel value 0-255). The methodology of the MRI brain human image classification is as follow:

- 1- Preprocessing Step using a median filter.
- 2- Feature extraction using Haar Wavelet and GLCM.
- 3- Feature selection by GA and K-NN.

- 4- Classification step using PNN algorithm. Block diagram of the proposed system is shown in figure 1.

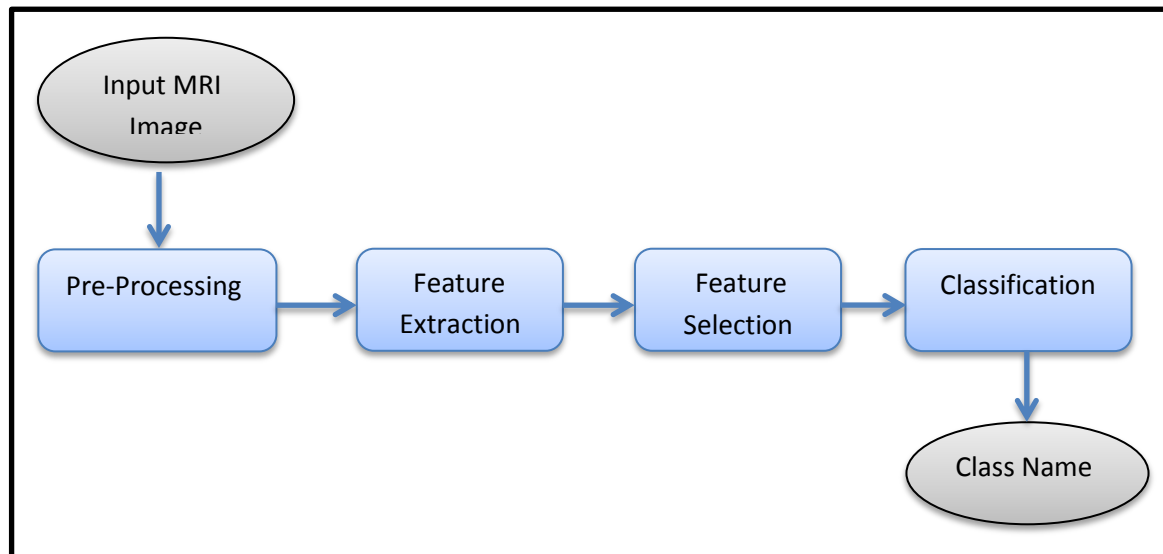


Figure 1: Block Diagram of the Proposed MRI Brain Tumor System.

A. Preprocessing Stage

In this step, we try to analysis the image which performs noise reduction and image enhancement techniques to enhance the image quality. In this step, we'll use a median filter.

▪ Median Filter

The median filter is used to reduce the salt and pepper noise present due to motion artifacts (movement of the patient during the scan) in the MRI images. It is done for smoothening of MRI brain image. Here we are using 3x3 (MRI) median filters to eliminate salt and pepper noise [6]. Figure 2 shows the after the applied median filter.

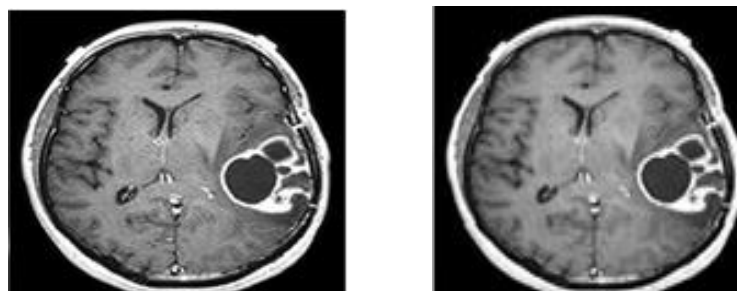


Figure 2: (a) input image and (b) After Applying Median Filter.

B. Feature Extraction Stage

Feature Extraction is a challenging task to extract a good feature set for classification. The purpose of feature extraction is to reduce the original data set by measuring features or certain properties, which

distinguish one input pattern from another. There are different feature extraction methods but in this section, the texture-based ones can be most effective for classifying the medical images. There are several texture-based feature extraction methods but Gray Level Co-occurrence Matrix (GLCM) is very common and successful [7]. In the proposed method one Level Discrete Wavelet transform (Haar Wavelet) is firstly used to decompose input image into four sub-images and then GLCM method is applied on each sub-image.

1. Discrete Wavelet Transform

The discrete wavelet transform is identical to a hierarchical sub-band system where the sub-bands are logarithmically spaced in frequency and represent octave-band decomposition. By applying DWT, the image is actually decomposed (i.e., divided) into four sub-bands in level one. As shown in figure 3 the critically sub-sampled of DWT [7]. As a result, there are four sub-band (LL, LH, HH, and HL) images at each scale. For feature extraction, only the four sub-bands are used for DWT decomposition at this scale then feature extraction based on GLCM.

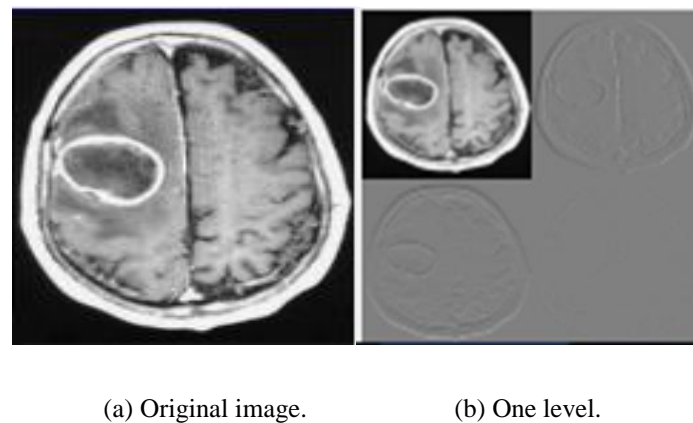


Figure 3: Image Decomposition.

2. Gray Level Co-occurrence Matrix (GLCM)

GLCM is used for feature extraction from MRI brain image. A feature of the image based on pixels and its neighboring pixels are extracted from image GLCM matrix is formed contains the textural feature based on two-pixel intensity values in the matrix. Feature-based on pixel and its neighboring pixel is extracted by GLCM (i, j) matrix. GLCM is a two-dimensional function, composed of n of horizontal direction pixels and m of vertical direction pixels. The horizontal and vertical coordinates of the image is given by i, j. $0 \leq i \leq n \leq j \leq m$ where total pixel number is $m \times n$. First, the intensity of the pixel and its neighboring pixel is calculated for the entire image. For getting more reliable texture feature multiple GLCMs are computed for different directions at (0° , 45° , 90° and 135°) which can give the spatial relationship between neighboring pixels[8]. This method reduces the computational complexity. After calculation for GLCMs of 4 sub-bands images, it is used to calculate features of the image which uniquely describes the images.

$$1. \text{Energy} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} P^2(i, j) \quad (2)$$

$$2. \text{Entropy} = - \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} P(i, j) \log_2 (P(i, j)) \quad (3)$$

$$3. \text{Contrast} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} P(i, j) * (i - j)^2 \quad (4)$$

$$4. \text{Homogeneity} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} \frac{P(i, j)}{1 + (i - j)^2} \quad (5)$$

$$5. \text{Variance (v)} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} (i - \mu_x)^2 p(i, j) \quad (6)$$

$$6. \text{Dissimilarity} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} |i - j| \times P(i, j) \quad (7)$$

$$7. \text{Maximum Probability} = \max \{ p(i, j) \} \quad (8)$$

$$8. \text{Sum Entropy (SE)} = - \sum_{i=2}^{2Ng} p_{x+y}(i) \log p_{x+y}(i) \quad (9)$$

$$9. \text{Sum Variance (SV)} = \sum_{i=2}^{2Ng} (i - SA)^2 p_{x+y}(i) \quad (10)$$

$$10. \text{Sum Average (SA)} = \sum_{i=2}^{2Ng} i p_{x+y}(i) \quad (11)$$

$$11. \text{Difference Variance (DV)} = \sum_{k=0}^{Ng-1} [k - \sum_{i=0}^{Ng-1} (i) p_{x+y}(i)]^2 p_{x+y}(i) \quad (12)$$

$$12. \text{Difference Entropy} = - \sum_{i=0}^{Ng-1} p_{x-y}(i) \log \{ p_{x-y}(i) \} \quad (13)$$

$$13. \text{Cluster Shade} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} [i + j - \mu_i - \mu_j]^3 p(i, j) \quad (14)$$

$$14. \text{Cluster Prominence} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} [i + j - \mu_i - \mu_j]^4 p(i, j) \quad (15)$$

$$15. \text{Autocorrelation} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} ij p(i, j) \quad (16)$$

$$16. \text{Inverse Difference Moment (IDM)} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} \frac{p(i, j)}{1 + (i - j)^2} \quad (17)$$

$$17. \text{Inverse Difference Normalized (IDN)} = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} \frac{p(i, j)}{1 + (i - j)} \quad (18)$$

$$18. \text{Correlation} = \frac{\sum_{i,j} [(ij) p(ij)] - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (19)$$

Where μ_x , μ_y , σ_x and σ_y are the means and standard deviations of p_x and p_y .

$$\mu_x = \sum_{i=0}^{Ng-1} i \sum_{j=0}^{Ng-1} p(i, j) \quad (20)$$

$$\mu_y = \sum_{j=0}^{Ng-1} j \sum_{i=0}^{Ng-1} p(i, j) \quad (21)$$

$$\sigma_x = \sum_a (a - \mu_x)^2 \sum_b p(a - b) \quad (22)$$

$$\sigma_y = \sum_b (b - \mu_y)^2 \sum_a p(a - b) \quad (23)$$

$$19. \text{Information Measure Correlation1 (IMC}_1) = \frac{H_{xy} - H_{xy1}}{\max\{H_x, H_y\}} \quad (24)$$

$$20. \text{Information Measure Correlation 2: (IMC}_2) = \sqrt{1 - \exp(-2(H_{xy2} - H_{xy}))} \quad (25)$$

Where, H_x & H_y are the entropies of p_x & p_y , respectively. While:

$$H_{xy1} = - \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} p(i, j) \log_2 [p(i) p(j)] \quad (26)$$

$$H_{xy2} = - \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} p(i) p(j) \log_2 [p(i) p(j)] \quad (27)$$

C. Feature Selection Stage

The medical image is a high volume in nature. If the data set contains redundant and irrelevant attributes, classification may produce a less accurate result. The genetic algorithm can deal with large search spaces efficiently, and hence has less chance to get a locally optimal solution than other algorithms. Our proposed algorithm consists of two parts:

- 1- The first part deals with evaluating features (chromosome) using a genetic algorithm (GA).
- 2- The second part deals with building classifier (K-NN) and measuring the accuracy of the classifier. In the proposed system, we used the KNN classifier with a different (K) at each time. Starting from $k=1$ to the $k = \text{the square root of the training set}$. Then our multi classifiers system uses majority rule to identify the class, i.e. the class with the highest number of votes (by 1-NN, 3-NN, 5-NN... \sqrt{n} -NN) is chosen.

1. K-Nearest Neighbor

This approach is one of the simplest and oldest methods used for pattern classification. It often yields efficient performance and, in certain cases, its accuracy is greater than state-of-the-art classifiers. The KNN classifier categorizes an unlabeled test example using the label of the majority of examples among its k-nearest (most similar) neighbors in the training set. The similarity depends on a specific distance metric; therefore, the performance of the classifier depends significantly on the distance metric used. The Euclidean distance between a test sample (x) and samples of a training set. For N-dimensional space, Euclidean distance between any two samples or vectors x and \bar{x} is given in (28) [14].

$$D = \sqrt{\sum_{i=1}^N (x_i - \bar{x}_i)^2} \quad (28)$$

2. Genetic Algorithm

A genetic algorithm is a general adaptive optimization search methodology based on a direct analogy to Darwinian natural selection and genetics in biological systems. GA work with a set of candidate solutions called a population. Based on the Darwinian principle of 'survival of the fittest', the GA obtains the optimal solution after a series of iterative computations. GA generates successive populations of alternate solutions that are represented by a chromosome, i.e. a solution to the problem until acceptable results are obtained. A fitness function assesses the quality of a solution in the evaluation step. As defined by formula (29).

$$\text{Fitness} = W_A \times \text{KNN_accuracy} + W_{nb}/N \quad (29)$$

Where W_A is the weight of accuracy, and it's can be set from (0.75 to 1). And W_{nb} is the weight of N features participated in classification where $N \neq 0$.

The crossover and mutation functions are the main operators that randomly impact the fitness value. Chromosomes are selected for reproduction by evaluating the fitness value. The fitter chromosomes have a higher probability to be selected into the recombination pool using the roulette wheel method. Crossover is a random mechanism for exchanging genes between two chromosomes using the one point crossover. In mutation the genes may occasionally be altered, i.e. in binary code genes changing genes code from 0 to 1 or vice versa. Offspring replaces the old population using the elitism or diversity replacement strategy and forms a new population in the next generation. Figure 4 illustrates the genetic operators of crossover and mutations [11] from the experiment the results were extracted, 10 features (energy, entropy, contrast, variance, sum entropy, difference entropy, homogeneity, cluster prominence, cluster shade, and dissimilarity) were selected from the set of 20 features and the population size, P was varied (50,100 and 500) .

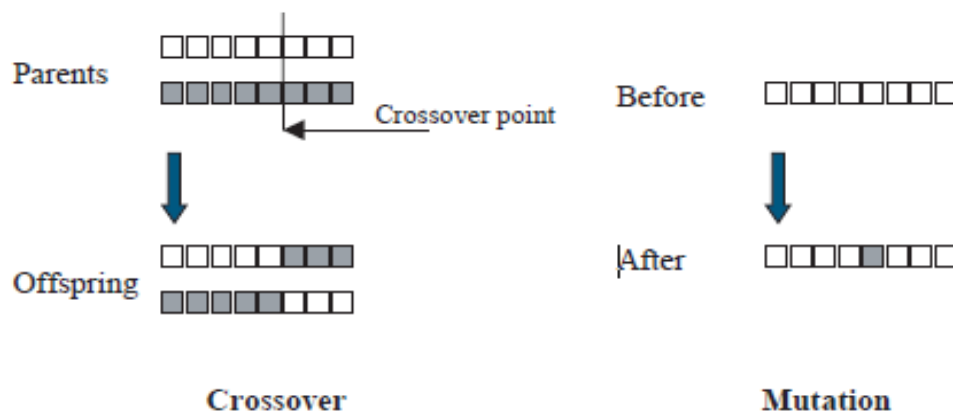


Figure 4: Genetic crossover and mutation operation.

3. Proposed Algorithm

Step (1): Input patterns of wavelet transform (M).

Step (2): Apply genetic search to generate the random population (G_i).

Step (3): Compute the transformed patterns (N) by applying the following equation

$$N = M \times G_i \quad (30)$$

Step (4): Calculates the accuracy of the classifier (K-NN) and returns to GA by the following equation

$$\text{Accuracy} = \frac{\text{no.of samples correctly classified in test data}}{\text{Total no.of samples in the test data}} \quad (31)$$

Step (5): Calculate the fitness value of the population by applying the function (29).

Step (6): Select the subset of higher fitness features.

Step (7): Crossover is done between the fittest individual.

Step (8): Mutation is done between the fittest individual.

Step (9): New population is created.

Step (10): If the generation is not ended, it will calculate fitness v value.

Step (11): End.

Figure 5 illustrates classification accuracy by using a GA-based features extractor.

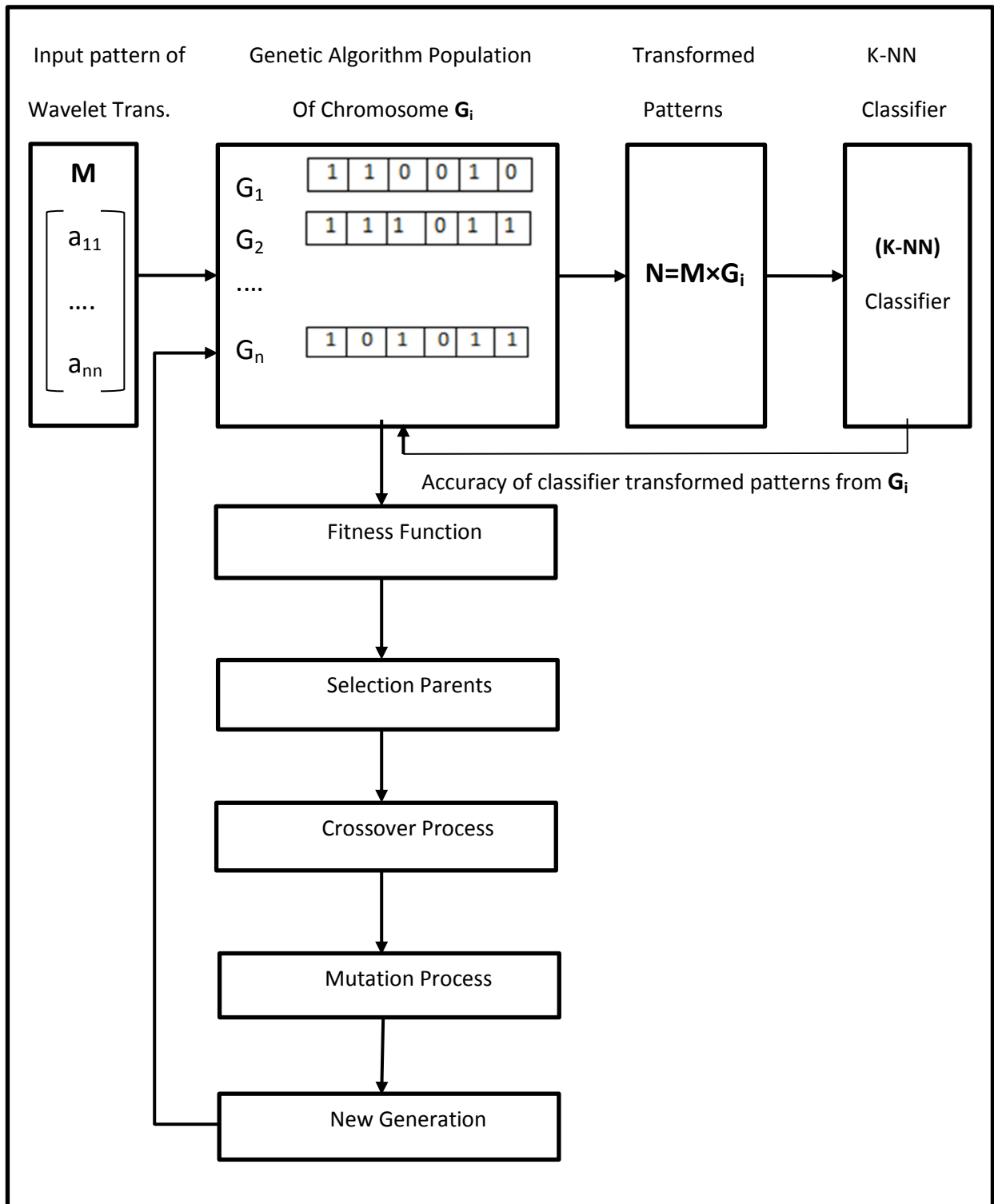


Figure 5: Classification Accuracy Using a GA-Based Features Extractor.

D. Classification Stage by Probabilistic Neural Network (PNN)

PNN is supervised feed-forward neural network algorithm derived from Bayes classifiers. It is used probability density function (pdf) for each class of training sample in the classification. If training sample is increased then classification goes near to density function of the class. The purpose of the probabilistic neural network (PNN) is to classification. In this stage, the test MRI brain image is compared with the training MRI brain image and gives output training MRI image which is similar to test image. PDF is given by the equation (32).

$$f_k(x) = \left(\frac{1}{(2\pi)^{\frac{d}{2}} \sigma^d} \right) \left(\frac{1}{N} \right) \sum_{i=1}^{N_k} \exp \left[- \frac{(x - x_{ki})^T (x - x_{ki})}{2\sigma^2} \right] \quad (32)$$

Where d = denotes the dimension of the pattern vector(x). i = pattern number, N = denotes the total number of samples in class, x_{ki} = vector of i -th training pattern from class 1, T = vector transpose. The σ is the " $\sigma_j = \text{STD}(X_i)$ " where X_i is the vector in training data and j number of classes. PNN algorithm consists of three layers. **The input layer** is the first layer which is the first distributed of the training input patterns. The number of neurons or nodes in the input layer is equal to the number of input vectors or variables. The second layer is **the pattern layer** or hidden layer. Each input vector in the training set has a processing element. Each element in the pattern layer is trained once. The third layer is **the output layer** (for each output class), an equal number of processing elements is used. Otherwise, the network will generate poor results. When an input vector matches the training vector, an element generates a high output value. Figure 6 illustrates the architecture of Probabilistic Neural Network [12].

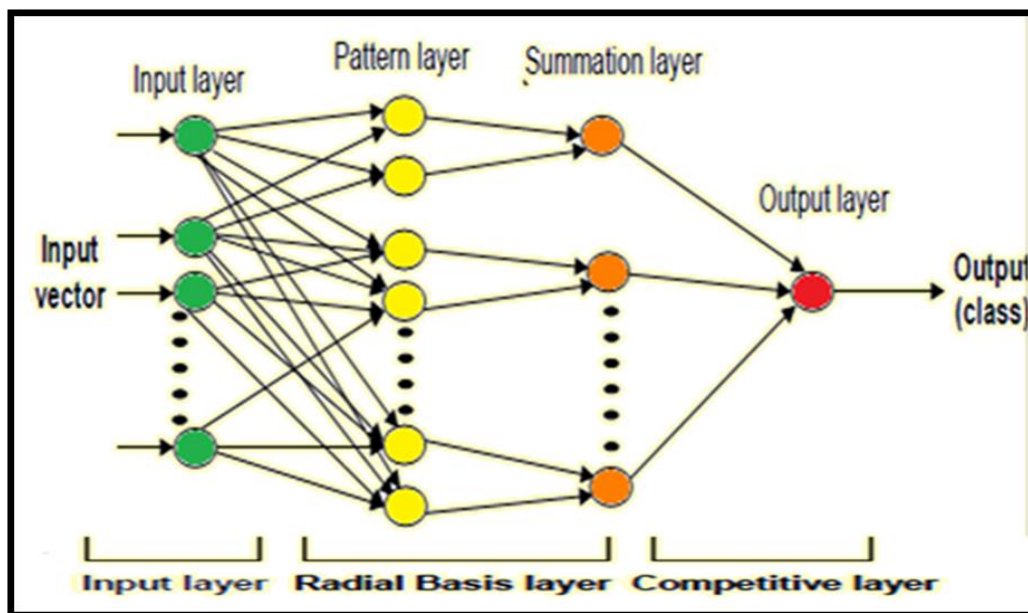


Figure 6: Architecture of Probabilistic Neural Network.

IV. RESULTS AND DISCUSSION

It includes the results of the classification of two systems and the comparison between them. In this paper, an automatic brain tumor classifier was proposed. The proposed technique was implemented on MRI dataset (these are Lymphoma, Cystic oligodendroglioma, Glioblastoma multiform, Meningioma, Ependymoma and Anaplastic astrocytoma). The numbers of collected images are 140. The algorithm described in this paper is developed and successfully trained in Visual Basic.Net.2013 using a combination of image processing and neural network toolbox. The remaining 70 MRI brain images from different types will be utilized as testing data phase. The result represents that 70 images are classified correctly. **The First System** is the classification of the MRI of the brain with 20 GLCM features (without a genetic algorithm). The classification rate of testing is 92.85%. **The second system** is (the proposed system) to classify the MRI images brain with 10 GLCM features using the genetic algorithm and K-NN. Classification rates of 4 cases (direction =0°, 45°, 90° and 135°) are 98.57 %, 100%, 97.14% and 98.57 % respectively. The maximum classification rate of testing is 100% in case=45° so the proposed system with a hybrid approach (Genetic Algorithm and K-NN classifier) is better than the first system (without Genetic Algorithm). Figure 7 illustrates a flowchart of the classification rate of the first system and the proposed system.

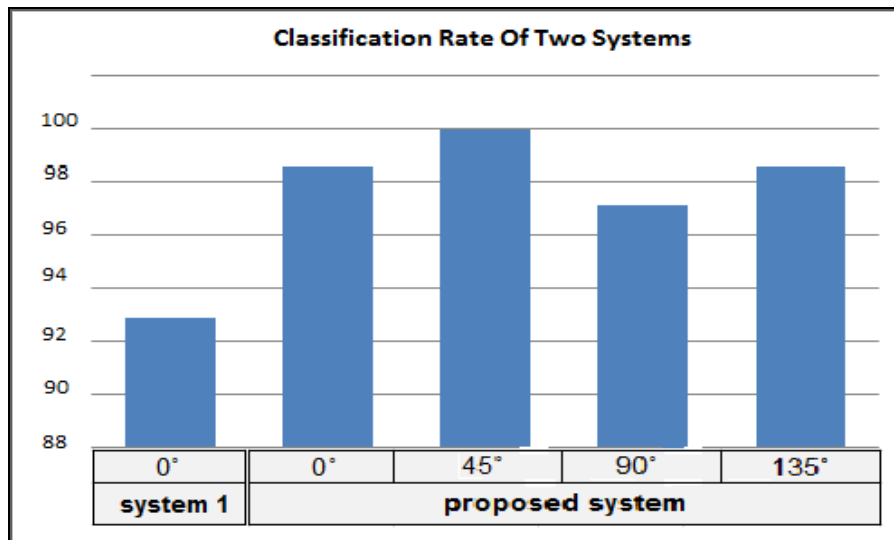


Figure 7: Flow Chart of the Classification Rate of the First System and the Proposed System.

The optimal solution of the proposed system is achieved with a population size of 100 (in GA) and classification results for the four bands is convergent ($L_1L_1=22\%$, $L_1H_1=26\%$, $H_1L_1=28$, and $H_1H_1=24$) as shown in figure 8. It is achieved an optimal solution in case ($k=7$) of the K-NN classifier.

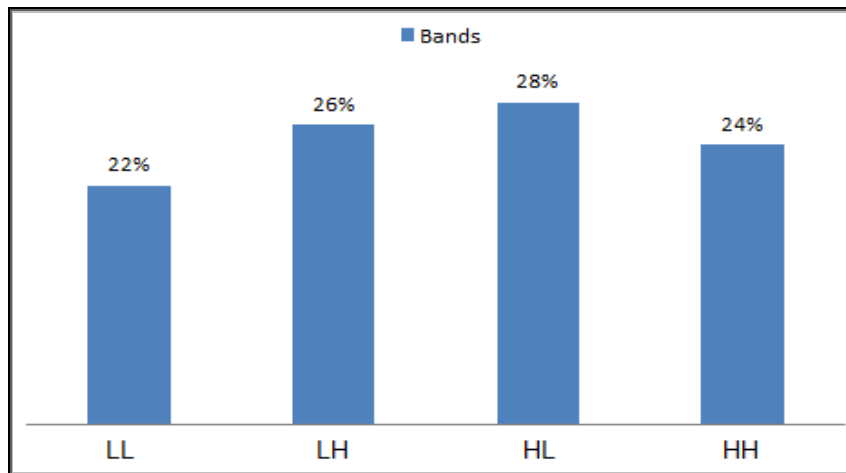


Figure 8: The Classification Rate for Four Bands

V. CONCLUSION

In this work, the new method is a combination of Discrete Wavelet Transform (Haar Wavelet), Genetic Algorithm, K-NN and Probabilistic Neural Network. By using this algorithm, an efficient Brain Tumor Classification method is been constructed with a maximum classification rate of 100 %.This method could serve inaccurate classification of Brain Tumor diagnosis.

VI. REFERENCES

- [1] Lashkari A., "A Neural Network-Based Method for Brain Abnormality Detection in MR Images Using Zernike Moments and Geometric Moments", International Journal of Computer Applications, Vol. 4, No. 7, pp. 1-8, July 2010.
- [2] Abdullah N., Ngah U., and Aziz S., "Image Classification of Brain MRI Using Support Vector Machine", Imaging Systems and Techniques (IST), IEEE International Conference, pp. 242-247, May 2011.
- [3] Ata'a A.H and Dhia A., "Classification Human Brain Images and Detection Suspicious Abnormal Area", IOSR Journal of Computer Engineering, Volume 18, May-Jun, 2016.
- [4] N. Pergad and K. Shingare, "Brain MRI Image Classification Using Probabilistic Neural Network and Tumor Detection Using Image Segmentation", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Vol. 4, Issue 6, 2015.
- [5] H. Naveena and K. Shreedhara, M. Rafi, "Detection and Classification of Brain Tumor using BPN and PNN Artificial Neural Network Algorithms", International Journal of Computer Science and Mobile Computing, Vol.4, Issue 4, India, 2015.

- [6] M. Sudharson, S.R. Thangadurai Rajapandiyam, and P.U. Ilavarasi, "Brain Tumor Detection by Image Processing Using MATLAB", Middle-East Journal of Scientific Research 24 (S1): 143-148, 2016.
- [7] Ali Sh., Mohammad R. A. and Mohammad H. N. Sh. "A CAD System for Automatic Classification of Brain Strokes in CT Images", International Journal of Mechatronics, Electrical and Computer Technology, ISSN: 2305-0543, pp. 67-85, Vol. 4(10), Jan 2014.
- [8] Dipanshu, N. Masalkar, and Shitole, A.S, "Advance Method for Brain Tumor Classification", International Journal on Recent and Innovation Trends in Computing and Communication vol.2, 2014.
- [9] Cheng L. and Chieh J., "A GA-based feature selection and parameters optimization for support vector machines", ELSEVIER, Expert Systems with Applications 31, 2006.
- [10] Ahmad B., Mohammad A., and Ahmad A., "Solving the Problem of the K Parameter in the KNN Classifier Using Ensemble Learning Approach", (IJCSIS) International Journal of Computer Science and Information Security, Vol. 12, No. 8, August 2014.
- [11] S.U Aswathy, G.Glan Devadhas and S.S.Kumar, "MRI Brain Tumor Segmentation Using Genetic Algorithm With SVM Classifier", Journal of Electronics and Communication Engineering, e-ISSN: 2278-2834, p-ISSN: 2278-8735 PP 22-26, 2017.
- [12] Swapnali S. and Dimple C., "Classification of Brain Tumor Using Discrete Wavelet Transform, Principal Component Analysis, and Probabilistic Neural Network", International Journal for Research in Emerging Science and Technology, Volume.1, Issue.6, pp.13-19, November 2014.

Information Retrieval thru Cellular Devices

Muhammad Zunnurain Hussain

Tahir Iqbal

Muhammad Zulkifl Hasan

Adeel. M

Sr. Lecturer CS & IT

Sr. Assistant Professor

Sr. Lecturer CS & IT

Student 6th Sem

Zunnurain.bulc@bahria.edu.pk

tahir.iqbal@bahria.edu.pk

zulkifl.hasan@lgu.edu.pk

Adeeladrees419@gmail.com

Bahria University Lahore Campus

Bahria University Lahore Campus

Lahore Garrison University Lahore.

LGU Lahore

Abstract

Information retrieval is the huge field and this research paper we are presenting the Information retrieval in which we explore relevance ranking using terms relevance using hyperlinks, synonyms, ontologies, indexing of documents web search engine directories and some other things there are some more difference sources that we use in Information retrieval let's talk about the mobile Information retrieval that article is Information retrieval such as design, moving things file folder, speech, voice, video, images and possible combination of mobile devices mobile it shall becoming even permanent in upcoming years according to previous studies mobile phone devices toward computer as a primary tool of internet. In Information retrieval we have three major examples physics recommended, e-commerce and movies and media sites the Information retrieval is a complex procedure Mainly there are three types of information retrieval. Information retrieval system knowledge based system database management system so in that case Information retrieval system we have web search engine keyword search of a extensive formula update capability is shared used for both read and write and legal system is significant detective capabilities and relevantly small schema.

Introduction

Retrieval of information is concerned with the sequences of files such as moving things design, speech voice, animation, pictures, images, voice, videos, and their combination that use in mobile phone devices and with the connectivity of wireless devices. The proliferation of mobile phones and other devices has created a large and huge amount of demands of information of mobile material well effective mobile information retrieval methods to see this need we shall require new methods and technologies for presenting, testing, modifying, and retrieval of the telephone data. The special features of mobile phone devices make them in different and more technique, and other way more initial and ancient, compared to its traditional counterparts. Mobile phone information retrieval is a subset of mobile information retrieval, as mobile information retrieval moves to the fore, two main feature character research in this new technologies sites area that is awareness about tourism and adapting content. In a broad sense, content adaption take fit the input of the mobile phone devices, and context awareness analyze the output from the mobile devices to the user, which can also be the fed back to the devices. The most of users still relay on

browse list maintained by mobile operators, search based search based search usage access the content that is moving fast, it's just like the transition that was to search engine in the initial web from directly services.

As the interface of mobile search terminals is designed as personal computers, commercial engines of other engines such as mobile, live search mobile, or a search are often a very painful and long-term, and consumers. Expensive experience for here is the article, we solve the problem of mobile search using search results cluster, which consists of organize the results achieve in response to a query into a hierarchy of labeled clusters is that reflect the various kind of components of the query topic. All through the queries train engine may not be of all type of factories it support some of the essential information where the simple search engine normally fails. The most wonderful feature of the hierarchy of feature is that it makes less or smaller way or something like shortcut way on similar subtopic items for an extraordinary question. If such things are placed correctly or with different clusters, the user may be able to choose the right path from the cluster label, these items can be

accessed linear time instead of linearism. We prepare and examine a detailed idea. Overall, the answer or result of our second experience has suggested that the search results were better than the listings of clustering effectively, across multiple devices, and also for some non-strict sub topic retrieval tasks. Clustering benefits and benefits of cluster on the listings were especially explicit on mobile phones. We also realized that the effectiveness of receiving list or cluster recovery can usually be reduced because search tools are small, although cluster can sometimes be better on smaller devices.

Compared to traditional things of the cluster, compared to the needs of many types of requirements, which can be picked up by the type of application in which it embedded, such as insight

Context Awareness Information Retrieval

Context aware computing a lot of things it is the notion of having computer system being aware of context of the word sense and it seem to be aware of the context means that you have system that can recognize that can proceed the real world use some sensors and can process upon those simulation which they get from the real world and so it just found very complicated in first place but it really simple idea what we do all the time so to navigate your way around you don't have to be familiar. There are some more applications in which we have agent that kind of application that take advantage of really processing the signals that are generated by a whole bunch of signals and making sense, there are different type of examples of context aware computing let's take an example we have a mobile if you look at sort of navigation system like navigation system of the phone this is context aware what is the context aware it is aware so obviously very simple it has GPS receiver in there that knows where it is and by this mean it guys and come back is moving with you. We have some more examples like traffic situation the time of day and perhaps even sort of what you like and what you dislike some people I'll try to avoid the motorway that on the small Street on the other way around so this is information this is contact information and if you have all this contact information and you want to say I want to go somewhere let's talk about the new example people have been in the front of the street light goes on it's basically a switch its look like if

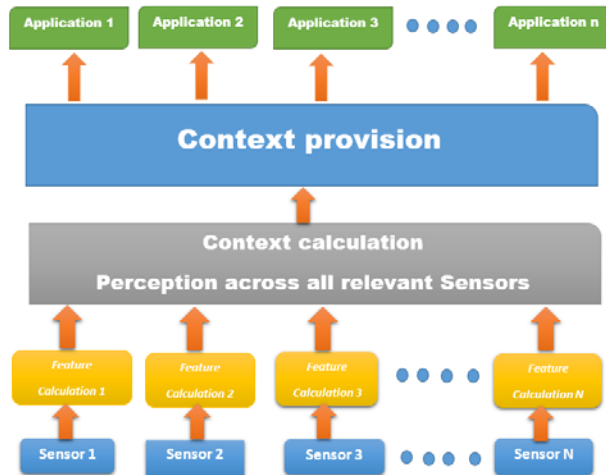
cluster labels, high computing performance, short Open input data, unauthorized number of clusters, and clusters. The use of clustering engine intended, the most important part is the quality of cluster labels, as opposed to improving the cluster structure only. The clustering algorithm has been developed due to the large variety of search results, the currently available clustering engine produces a lot of different clusters and cluster labels for the same question. There are three high-definition lines available. First of all, many web cluster engines are recently offered to search for a desktop, the second is a CREDO clustering engine, the reference system used in this work, about the last one alternative mobile information recovery technique. Is. They are discussed, in the next section.

somebody moving to switches on the light it's come the tricky part.

Context awareness is a central enabling Technologies for your vehicle is computing and if you look back at the beginning of the 90s when mark why sir published his article on computing the work the research of XEROX PARC it was very much this notion of sort of have world where sort of computers are everywhere and now we mention we have to sort of do what you have thought of all the configuration all the setting up each of those computers of hundreds of computers around this is not going to school because you need to take some data that store your computer and tell them what they have to do and so it's very clear if you have a lot of computers they must work together and they must be adaptive to that to the real world and there is a very interesting quotes from Mark Vizag about Technologies believe themselves into everyday life until they interesting miserable from the Technologies can become invisible that's mean you can't anything but it is too little that you see really realize that this technology is there you just use it and most people who are nowadays using machines and gadgets there are some processor that just 4 or 8 bit processor and the interesting points of Computer has sort of move into everyday life and think half see the context awareness is a central point to make things invisible so that we don't have to take care of all those things work automatically but as we said earlier it really are tricky part the computer thing will wrong.

The context awareness at beginning of 90's the first paper were published so there was obviously work before and think the notion of context awareness is sort of (Bill Cheliya) paper from 1994 and his

dissertation in 90's that's really sort of laid a foundation of concepts of computing and he define a notion of lot of route dated at that time what about the research and location based finding of resources but the idea of contexts that was very much described in it that the paper in 1994 even so it has only 4 pages. If people read that one they probably would not write a lot of other paper because most of the ideas about context awareness or in that paper already and so it is a sort of the mandatory did (Spill WM-CSE) in 1994.



Content Adoption

It is important to focus on research efficient way to execute mass data and small screens we need both efficient and efficient algorithm that can perform well in Limited electrical setting it display some research topic the area many of these areas can help in adopting current IR Information retrieval Technologies for mobile data such as content original indexing concerned and mobile data recovery in addition tradition and Information retrieval Technologies can be adopted for search of knowledge through mobile summary index and table of content and keywords in retrieval. As the mobile data is growing we need a scaling browsing algorithm for efficient recovery models and query

Approaches of Mobile Information Retrieval

Some approaches of mobile Information retrieval is machine learning and content of Information retrieval take care of query suggestion and take care of interface design question suggestion is care using ontologies profile mining and awareness about

The starting point of context awareness and then with the availability of broiler fences special GPS more work toward the end of the 90s and 2000 reader then sort of there was a project of Manchester University that guy project in 1997 nothing 1998 were people experimented with tablet species of very large at that time and not very powerful but having them a sort of personal comic books electronics books around the city and table contacts of her and they try to go beyond the location looking at condition weather that your personal profile at opening hours. Before showing quality aspects of context and services, the modeling of situation data is figured out. Both are compulsory to define, handle and analyze the environment and its changes, to react sensitive to it. How context and adapting services get send and received, is mentioned in the section of dividing, followed by a short overview of current researches and services, to view some achievement of context-aware software and how development tools can ease the process of construction. Finally, this paper ends with a short overview and a look into the future.

processing of mobile information as well as large mobile database. Although there are many topics related to detailed research on every broad topic, the integrated mob and content adaptation of mobile IR's new frontier contexts will be combined. Ideal technology should marry the past to make a new foundation for the mobile. However, this is a very high level of special treatment. Traditionally there is no source code, but the developer has established user interface and data processing units in the development environment. Program flow is a well-structured. Today, the main heritage of the mission plays an important role in maintaining the software.

context Technologies can be adopted for search of knowledge through mobile summary index and table of contents in Information retrieval on the other hand, the actual information on content delivery techniques is used for mobile platforms The main layout design is responsible for the screen size and resolution, input type etc., whereas the interaction the design is responsible for the inputs that are given like all keywords given.

Mobile User Interface

Our approach also faces these problems and the previous results are to be met, but lessons are learned and in the purpose of literature available, facilitate the prototype of foreign and prototype in facilitating the user's involvement and design process. The purpose of integrating the diagnostic phase is to aim. None of the current work includes antivirus prototype and evaluation on real devices, including intuitive and simple for adjusting prototype or analyzing individual analysis, or simultaneously. In addition, focuses on the most specific issues or domains and the server depends on client architecture, which is consistently connected or repeatedly. In addition, no different techniques are combined in any device, allowing designers to evaluate their needs. Our goal is to focus on that which is directly related to prototypes on the initial design stages, simplifying the analysis facilities at its place through many mechanisms and easily collected data by end users. To combine For this, diagnostic procedures automatically recognize two stages that are integrated within mixed-sized prototype tool, flip space between prototype and internal condition.

Our previous collaboration during the previous work is integration of many data collection and analysis techniques that can be used indefinitely on

Content Delivery Techniques

The needs and preferences of some different information should be notified about mobile devices. They want to know their interests shown in the liked format. For example, in terms of knowledge interest, some consumers are interested in online shopping and want to get information about Discounts or promotion about clothes or electronics products, while other users of the NBA big fans And want to get the latest NBA latest scores.

For preferred formats, some users prefer text or video clips besides text messages, although it can cause a long delay. However, other users just want text messages to reduce download delay. Another example is that job-makers may need a step by

Query Suggestion

We analyzed trials with "hard coded" tips of twenty five such users who saw suggestions on studying how the movement influences acceptance patterns. The remaining four hundred fifty cases were

mobile devices, to facilitate user participation and design process. The meaning of the basic pattern is described as a loop learning language, which has three main functions: search representation, encouraging them, and transit them. Finally, information set with non-stable schemes is used to complete overall emotional work. Identification of two key types of reliable search engines, "an off" work and recurring tasks were identified. For an off-task, the aim of the mentor is to take advantage of the process better. Recurring tasks, purpose is to improve the benefits of work cycle again. Juice and colleagues say that in many cases, maximum expenses are in time to spend, time spent, and to find, search and select more relevant information.

In addition, the main character of the representative design has been identified. In the context of the Web Search Interface, it appears that the results are to be provided through a proper representation and one representation of the change, and support of the origin of data. For example, there is a change when there is a list of a result of the resulting selection and switching between one type of review, which results in different levels of results.

defining an application's step-by-step job, but the customers who are familiar with the application may have at least been able to complete their tasks Shortcuts (only one or two pages of time). Let's talk about content delivery techniques there are a couple of examples basically content delivery techniques what kind of buffer between your website and anyone trying to access it so what you do is too kind to reduction your DNS records through the content delivery techniques so people who are accessing your website don't necessary directly interact with your website Sara the first have to go through this content delivery techniques and there are lot of benefits for this not just for security but also performance so let's gets into those as well first let's talked about how this would improve performance

considered for time and critical press analysis. When evaluating the average time to enter a question, we exposed 44 questions where the user either used the key or entered the question

incorrectly and showed the error screen. Among those 450 questions entered on the interface, which are included in the drop-down list of suggestions, 435, showed that useful suggestion presented before the user entered the question. We consider partial completion of useful tips where the proposal completes the portion of the desired required question, most complete where the required keyword or the complete completion super net is where the advised is the required question. The distribution of useful tips was extended towards complete completion; 348 questions were displayed in the list of their suggestions with complete completion, fully displayed. As far as mentioned, figures in this paper only refer to complete

Measuring Data transfer Energy

In order to cope with energy consumption of network activities inspired by the IR, we first need to deal with the energy consumption of individual data transfer. For this end, we get a measurement of data measurement for a wide range of data size. Ideally, we want to evaluate the rest of the battery life and after the transfer of each data to determine the energy consumption. However, it is difficult to accurately present the remaining battery life. For example, the device used in this study provides measurement of remaining battery life in the measurement of remaining weight and is not very reliable due to battery drain non-linear. Instead, we use the following procedure: From the full charged state, we are looking forward to reaching every 10 seconds of transferring, as long as the battery is completely snack. Based on the measurement of battery life and average transfer time, we estimate the average energy of database transfer as shown below.

Smart Message Service (SMS)

SMS and MMS is a great messaging service in which Smart Mobile phones are a great way to exchange a large information for all mobile users. Service message service has increased from the marketing company, communication of all types of messages on the phone. There are many SMS developers available to control and prove the legal status of SMS, but the authors present in SMS Controller on an all-based basis, in which the previous features include content based SM S

completion. Histogram of a letter number before complete completion in the suggestions. Three questions whose users were selected from Google's login were selected and the following 4 interruptions were met. Each question included only letters and vacations, length of fifteen twenty, which included free space, requires a thirty-key press that does not use multi-tap inputs, tips, and any errors. , And there are two letters that were on the same key. Nine key keypad for each question, the length of the key and the length of the key pressure was according to the average length of mobile questions and the average number of key pressure required to enter.

Detection, Group Chat, SMS Text Analysis, etc. Progress and exciting new features and auto reply.

Smart Shopping

Recently, purchases have increased via mobile apps. With real-based mobile applications, the real world environment has the ability to play virtualized environments. Smart shopping mobile app meets the image that embedded in a special angle with the embedded mobile camera, with the image in its database and output details. This app is very easy for busy buyers, however the storage and recovery of photo details is still a challenge. This challenge has been made possible by the Internet using advanced image processing technology and web service app. Today's mobile apps are used in many aspects of life. In addition to Columns and SMS, business customers are using apps to improve customer satisfaction through smart shopping. However, there are many issues, such as analyzing customer behavior patterns / behavior, and how customers can provide the latest information.

Prepares a system that uses NFC, mobile and web application to provide users with the latest information at the same time. will help solve the problem in collecting and analyzing consumer purchasing habits. . This app is great, but it has some limitations and complexity. Users' mobile phones should have an NFC enabled Android smartphone, works online only and lacks security and privacy. The latest trends in mobile use have seen increasing the growth of mobile marketing. With the help of social vector and RFID technology customers were

offered to provide them with a smart shopping experience. There are many challenges in this development to identify how customers are and how they can be custom marketing messages. Generate and shipped to customers. Meanwhile, there are three perspectives to combat these challenges. The basic features of social vector,

Opportunities

Better results can be achieved if a user is measured by the quality of quality of quality, its device, which supports the measurement and analysis of the accurate, organized, and mobile-quality quality of education analysis. Real video and demand and VoIP apps were measured with more than 80% accuracy, which is close to or above the domain offered by experts. Especially in the medical field the use of natural language for the mobile app is increasing. We are talking about some of the more common jobs rules that you will see listen and what they Intel so when it come to information Technology you know certainly this a lot of different rules write their software programmer and write the software write the code for what we do their hardware engineer who are specialization in memory or monitor or hard disc type and then there engineers that works in networking technology which typically fall into one of about 4 different categories make sure you understand it these jobs descriptions are very fluid in other words what a person doing one of these jobs will actually do depends on the business of the company the work for example if you work for a really small company you know let's say you work for a company of 100 people of less and you are the charge of the network. Similarly, the prototype platform console of mobile applications is an open source prototype platform for mobile applications compatible with mobile applications. These devices and applications can also be used in emergencies and personal family situations, where the device can be worn by drivers and shows the sensor. Mobile purchase is due to the development of many types of apps. Some social vectors and RFID technologies are used, while others are designed to collect consumer purchases such as users, Wi-Fi, NFC and web services and share information.

governance based approach and comparative approach. Some limitations that are included in this app, besides a smartphone, the customer needs to take a smart card, the smartphone and the app registering the app need to download. Also can be saved on the individual server's server.

Integrated Features

Smartphone technology has been improved the use of mobile phones in this world has been rejected. Integrated features can be smartly used in case of WIFI, GPS navigation, HD cameras, touch screen and Internet access information and personal safety issues. He has proposed a mobile app that will make a decision based on driver's heart rate, driver's location and mobile phone sensor Integrated features of information refrigeration that we are trying to present, where the problem was resolved shortly to resolve the issue, which is facing the heart's problems and defines some of its axis The way to do There are some essential features that you need to fulfill each of your experiences and this is the process we can't hear. With more than Vector and RFID, NFC, GPP and higher resolution cameras and web applications., a more realistic app can be designed for mobile shopping. Shopping outlets can put their best prices in the database and apps can scare their prices and compare between other shops and make a report for customer according to context. Integrated experience want to brief indicate there is that Shad experience you're having now let's say in the color is similarity and other you who is experienced the same thing without the colors and another one Without seeing it other than places and sounds, anyone else has a great experience on just one left and on a small part of the neutral experience that you value as well. In addition, with Lou Gates, mobile app developers and providers can easily log into events from iOS, Android and HTML5 apps, and can quickly show this information to backup application requests. Meet events, operating systems and events in infrastructure.

Conclusion

In this article, we discuss web-information methods and tools that take advantage of web features, to reduce some of the problems that are the result of web recovery. In order to evaluate the information of the recovery of information, we used diagnostic measures like ancient and memory and studied how to count them effectively. Since the degree of effect depends on the efforts of many consumers, we have discussed about dealing with user efforts using the function of DCG and discount (discount discounted family).

Impact evaluation is an important aspect of investigating and designing information systems. More research has been made on the topic, and every day continues every year. The matter of decision-making and assessment of cost-effective compatibility is important. Interestingly, the interest in promoting user models for the individual, independent document compatibility has increased recently, Consolidation of continuous work in innovation and diversity, between the unmatched information related to documentation and documentation is investigating.

References

1. J.Q. Anderson and L. Rainie, "The Future of the Internet III, "Pew Internet and American Life Project, 2008; www.pewinternet.org/Reports/2008/The-Future-of-the-Internet-III.aspx.
2. T. Claburn, "Google CEO Imagines Era of Mobile Supercomputers," *Information Week*, 28 Oct. 2009; www.informationweek.com/news/software/web_services/showArticle.jhtml?articleID=220900806.
3. T. Tezuka, T. Kurashima, and K. Tanaka, "Toward Tighter Integration of Web Search with a Geographic Information System," *Proc. 15th Int'l Conf. World Wide Web (WWW 06)*, ACM Press, 2006, pp. 277–286.
4. A.T. KWEE and F.S. Tsai, "Mobile Novelty Mining," *Int'l J. Advanced Pervasive and Ubiquitous Computing*, to be published, 2010.
5. M. Aliannejadi, D. Rafailidis, and F. Crestani. Personalized Keyword Boosting for Venue Suggestion Based on Multiple LBSNs. In *Proceedings of the 39th European Conference on Information Retrieval*, pages 291–303, Aberdeen, Scotland, UK, 2017.
6. J. Allan, B. Croft, A. Moffat, and M. Sanderson. Frontiers, Challenges, and Opportunities for Information Retrieval: Report from SWIRL 2012 the Second Strategic Workshop on Information Retrieval in Lorne. *SIGIR Forum*, 46(1):2–32, May 2012.
7. T. A. Almeida, J. M. G. Hidalgo, and A. Yamakami. Contributions to the Study of SMS Spam Filtering: New Collection and Results. In *Proceedings of the 11th ACM Symposium on Document Engineering, DocEng '11*, pages 259–262, New York, NY, USA, 2011. ACM.
8. O. Arkhipova and L. Grauer. Evaluating Mobile Web Search Performance by Taking Good Abandonment into Account. In *Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '14*, pages 1043–1046, New York, NY, USA, 2014. ACM.
9. K. Berberich, A. C. König, D. Lymberopoulos, and P. Zhao. Improving Local Search Ranking Through External Logs. In *Proceedings of the 34th International ACM SIGIR Conference on*

Research and Development in Information Retrieval, SIGIR '11, pages 785–794, New York, NY, USA, 2011. ACM.

10. P. Borlund. The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. *Information Research*, 8(3), 2003.
11. O. Boudighaghen, L. Tamine-Lechani, and M. Boughanem. Dynamically Personalizing Search Results for Mobile Users. In *Flexible Query Answering Systems: 8th International Conference, FQAS 2009, Roskilde, Denmark, October 26–28, 2009. Proceedings*, pages 99–110, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.
12. B. A. T. Brown, F. Bentley, S. Bakhshi, and D. A. Shamma. Ephemeral Photo work Understanding the Mobile Social Photography Ecosystem. In *Proceedings of the Tenth International Conference on Web and Social Media, Cologne, Germany, May 17–20, 2016*, pages 551–554, 2016.
13. C. Carpineto, S. Mizzaro, G. Romano, and M. Snidero. Mobile Information Retrieval with Search Results Clustering: Prototypes and Evaluations. *Journal of the American Society for Information Science and Technology*, 60(5):877–895, May 2009.
14. E. Chang, F. Seide, H. M. Meng, Z. Chen, Y. Shi, and Y.-C. Li. A system for spoken query information retrieval on mobile devices. *IEEE Transactions on Speech and Audio Processing*, 10(8):531–541, Nov 2002.
15. L. Chen and L. Qi. A Diary Study of Understanding Contextual Information Needs During Leisure Traveling. In *Proceedings of the Third Symposium on Information Interaction in Context, IliX '10*, pages 265–270, New York, NY, USA, 2010. ACM.
16. P. Coppola, V. Della Mea, L. Di Gaspero, D. Menegon, D. Mischis, S. Mizzaro, I. Scagnetto, and L. Vassena. The Context-Aware Browser. *IEEE Intelligent Systems*, 25(1):38–47, Jan 2010.
17. G. V. Cormack, C. R. Palmer, and C. L. A. Clarke. Efficient Construction of Large Test Collections. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '98*, pages 282–289, New York, NY, USA, 1998. ACM.
18. F. Crestani. Vocal Access to a Newspaper Archive: Assessing the Limitations of Current Voice Information Access Technology. *Journal of Intelligent Information Systems*, 20(2):161–180, 2003.
19. F. Crestani and H. Du. Written Versus Spoken Queries: A Qualitative and Quantitative Comparative
20. Analysis. *Journal of the American Society for Information Science and Technology*, 57(7):881–890, May 2006.
21. F. Crestani, M. Dunlop, M. Jones, S. Jones, and S. Mizzaro. Theme issue on interactive mobile information access. *Personal and Ubiquitous Computing*, 10(4):193–194, 2006.

A SECURITY OVERVIEW OF WIRELESS SENSOR NETWORK

Muhammad Zulkifl Hasan¹, Zaka Ullah², Muhammad Zunnurain Hussain³

^[1,2]Department of Computer Science

^[1,2]Lahore Garrison University, Lahore, Pakistan, ³Bahria University Lahore Campus, Lahore Pakistan

zulkifl.hasan@lgu.edu.pk, zakaullah@lgu.edu.pk, zunnurain.bulc@bahria.edu.pk

Abstract

As of late, remote sensor organize (WSN) is utilized in numerous application zones, for example, checking, following, and controlling. For some utilizations of WSN, security is an essential necessity. In any case, security arrangements in WSN vary from conventional systems because of asset confinement and computational requirements. This paper investigates security arrangements: Tiny Sec, IEEE 802.15.4, Twists, Mini SEC, LSec, LLSP, LISA, and Drawl in WSN. The paper additionally introduces qualities, security prerequisites, assaults, encryption calculations, and operation modes. This paper is thought to be valuable for security planners in WSNs.

Keywords: Tiny Sec, IEEE 802.15.4, Twists, Mini SEC, LSec, LLSP, LISA, and Drawl in WSN.

Introduction

A wireless sensor network consists of substantial number of distributed nodes which works to gather information for making a decision about a logical problem or issue. There comes a situation where information security is a critical issue. Data loss or integrity is questionable in some scenarios which led to massive loss to applications which are critical in context to military, home security & in health sector. Sensor networks are prone to several traditional attacks with addition to newly introduced attacks such as, Sinkholes. The main aim of any attack is to interrupt the functioning of existing network with modification in data as an unauthorized user in the network. In this paper, we will critically discuss the security issues & handling with all types of attacks.(Abd-El-Barr, Al-Otaibi, & Youssef, 2005)

Challenges in Wireless Sensor Networks

Few critical factors in WSNs which are considered as challenges for the use of network are:

Power Constraints: Most of the applications of sensor network requires more power to operate but due to limited power many applications fail.

Limited Resources: Sensor node size is a critical factor for allocating bandwidth & computation abilities. The smaller the size of node the more are the limitation.

Flexibility & Portability of nodes: Due to frequent movement of nodes in sensor network it encounters situation where link failure changes the nodes communication throughout the network. (Anjali, Shikha, & Sharma, 2014)

Related Work

Sweeping examination is being done in the region of Remote Sensor Systems. Analysts have been focusing on settling an assortment of difficulties extending from constrained asset capacities to secure correspondence. Writing demonstrates that sensor systems are conveyed in broad daylight or relinquished zones, over unreliable remote channels, it is along these lines charming for a pernicious gadget/interloper to spy or infuse messages into the system. The customary answer for this issue has been to take up procedures, for example, message confirmation codes, open key cryptography and symmetric key encryption plans. In any case, since there are asset shortages for bits, the real test is to devise these encryption strategies in an effective way without yielding their rare assets. One strategy for protecting any system against outer assaults is to apply a direct key foundation. Be that as it may, it is realized that worldwide keys don't give arrange flexibility and match shrewd keys are most certainly not hearty arrangement. A more instinctive arrangement is required for WSNs(Bin Lu, Habetler, Harley, & Gutierrez, 2005).

Security Requirements in WSNs

It imparts a few shared traits to a run on the mill PC organize, yet in addition shows numerous qualities which are remarkable to it. The security

benefits in a WSN ought to ensure the data conveyed over the system and the assets from assaults and misconduct of hubs. The most vital security necessities in WSN are recorded underneath:

Information privacy: The security instrument ought to guarantee that no message in the system is comprehended by anybody aside from expected beneficiary. a sensor hub ought not enable its readings to be gotten to by its neighbors unless they are approved to do as such, key dispersion component ought to be to a great degree hearty, open data, for example, sensor characters, and open keys on the hubs ought to likewise be encoded in specific bodies of evidence to secure against movement investigation assaults.

Information trustworthiness: The component ought to guarantee that no message can be adjusted by an element as it navigates from the sender to the beneficiary.

Accessibility: These prerequisites guarantees that the administrations on a WSN ought to be accessible constantly even in nearness on an inside or outer assaults, for example, a dissent on administration assault. While a few instruments make utilization on extra correspondence among hubs, others propose utilization on a local access control framework to guarantee effective conveyance on each message to its beneficiary.

Security vulnerabilities of WSN:

Remote Sensor Systems are powerless against different sorts of assaults. These assaults are for the most part of three sorts. Assaults on mystery and confirmation: standard cryptographic procedures can secure the mystery and legitimacy of correspondence channels from pariah assaults for example, listening stealthily, parcel replay assaults, and adjustment or mocking of parcels. Assaults on arrange accessibility: assaults on accessibility of WSN are regularly alluded to as dissent of-benefit (DoS) assaults. Stealthy assault against benefit respectability: in a stealthy assault, the objective of the assailant is to influence the system to acknowledge a false information esteem. For instance, an assailant bargains a sensor hub and infuses a false information esteem through that sensor hub. In these assaults, keeping the sensor arrange accessible for its proposed utilize is basic. DoS assaults against WSNs may allow genuine

harm to the wellbeing and security of individuals. The DoS assault more often than not alludes to an enemy's endeavor to upset, subvert, or crush a system. Be that as it may, a DoS assault can be any occasion that reduces or disposes of a system's ability to play out its normal capacities (Dr. G. Padmavathi, 2009).

Feasibility of Basic Security Schemes in Wireless Sensor Networks

Security is a comprehensively utilized term including the qualities of verification, honesty, protection, nonrepudiation, and hostile to playback. The more the dependency on the information provided by the networks has been increased, the more the risk of secure transmission of information over the networks has increased. For the secure transmission of various types of information over networks, several cryptographic, stenographic and other techniques are used which are well known. In this section, we discuss the network security fundamentals and how the techniques are meant for wireless sensor networks.

Cryptography

The encryption-decryption techniques devised for the traditional wired networks are not feasible to be applied directly for the wireless networks and in particular for wireless sensor networks. WSNs consist of tiny sensors which really suffer from the lack of processing, memory and battery power/ Applying any encryption scheme requires transmission of extra bits, hence extra processing, memory and battery power which are very important resources for the sensors' longevity. Applying the security mechanisms such as encryption could also increase delay, jitter and packet loss in wireless sensor networks. Moreover, some critical questions arise when applying encryption schemes to WSNs like, how the keys are generated or disseminated. How the keys are managed, revoked, assigned to a new sensor added to the network or renewed for ensuring robust security for the network. As minimal (or no) human interaction for the sensors, is a fundamental feature of wireless sensor networks, it becomes an important issue how the keys could be modified time to time for encryption. Adoption of pre-loaded keys or embedded keys could not be an efficient solution.

Public Key Cryptography: Average Energy Costs of Digital Signature and Key Exchange in Millijoules (mJ)

| Algorithm | Signature | | Key Exchange | |
|-----------|-----------|--------|--------------|--------|
| | Sign | Verify | Client | Server |
| RSA-1024 | 304 | 11.9 | 15.4 | 304 |
| ECDSA-160 | 22.82 | 45.09 | 22.3 | 22.3 |
| RSA-2048 | 2302.7 | 53.7 | 57.2 | 2302.7 |
| ECDSA-224 | 61.54 | 121.98 | 60.4 | 60.4 |

Steganography

While cryptography aims at hiding the content of a message, steganography aims at hiding the existence of the message. Steganography is the art of covert communication by embedding a message into the multimedia data (image, sound, video, etc.)). The main objective of steganography is to modify the carrier in a way that is not perceptible and hence, it looks just like ordinary. It hides the existence of the covert channel, and furthermore, in the case that we want to send a secret data without sender information or when we want to distribute secret data publicly, it is very useful. However, securing wireless sensor networks is not directly related to steganography and processing multimedia data (like audio, video) with the inadequate resources of the sensors is difficult and an open research issue.

Physical Layer Secure Access

Physical layer secure access in wireless sensor networks could be provided by using frequency hopping. A dynamic combination of the parameters like hopping set (available frequencies for hopping), dwell time (time interval per hop) and hopping pattern (the sequence in which the frequencies from the available hopping set is used) could be used with a little expense of memory, processing and energy resources. Important points in physical layer secure access are the efficient design so that the hopping sequence is modified in less time than is required to discover it and for employing this both the sender and receiver should maintain a synchronized clock. A scheme as proposed in could also be utilized which introduces secure physical layer access employing the singular vectors with the channel synthesized modulation.

Security Threats and Issues in Wireless Sensor Networks

Most of the threats and attacks against security in wireless networks are almost similar to their wired counterparts while some are exacerbated with the inclusion of wireless connectivity. In fact, wireless networks are usually more vulnerable to various security threats as the unguided transmission medium is more susceptible to security attacks than those of the guided transmission medium. The broadcast nature of the wireless communication is a simple candidate for eavesdropping. In most of the cases various security issues and threats related to those we consider for wireless ad hoc networks are also applicable for wireless sensor networks. These issues are well-enumerated in some past researches and also a number of security schemes are already being proposed to fight against them. However, the security mechanisms devised for

wireless ad hoc networks could not be applied directly for wireless sensor networks because of the architectural disparity of the two networks. While ad hoc networks are self-organizing, dynamic topology, peer to peer networks formed by a collection of mobile nodes and the centralized entity is absent the wireless sensor networks could have a command node or a base station (centralized entity, sometimes termed as sink). The architectural aspect of wireless sensor network could make the employment of a security schemes little bit easier as the base stations or the centralized entities could be used extensively in this case. Nevertheless, the major challenge is induced by the constraint of resources of the tiny sensors. In many cases, sensors are expected to be deployed arbitrarily in the enemy territory (especially in military reconnaissance scenario) or over dangerous or hazardous areas. Therefore, even if the base station (sink) resides in the friendly or safe area, the sensor nodes need to be protected from being compromised.

Attacks in Wireless Sensor Networks

Attacks against wireless sensor networks could be broadly considered from two different levels of views. One is the attack against the security mechanisms and another is against the basic mechanisms (like routing mechanisms). Here we point out the major attacks in wireless sensor networks

Denial of Service

Denial of Service (DoS) is produced by the unintentional failure of nodes or malicious action. The simplest DoS attack tries to exhaust the resources available to the victim node, by sending extra unnecessary packets and thus prevents legitimate network users from accessing services or resources to which they are entitled. DoS attack is meant not only for the adversary's attempt to subvert, disrupt, or destroy a network, but also for any event that diminishes a network's capability to provide a service. In wireless sensor networks, several types of DoS attacks in different layers might be performed. At physical layer the DoS attacks could be jamming and tampering, at link layer, collision, exhaustion, unfairness, at network layer, neglect and greed, homing, misdirection, black holes and at transport layer this attack could be performed by malicious flooding and desynchronization. The mechanisms to prevent DoS attacks include payment for network resources, pushback, strong authentication and identification of traffic.

Attacks on Information in transit

Sensors in sensor organize watch the varieties of different factors or qualities and state to the sink as

indicated by the necessity. While sending the report, the data in travel might be adjusted, mock, replayed again or vanished. As remote correspondence is powerless against listening stealthily, any aggressor can screen the movement stream and get vigorously to intrude on, block, alter or create parcels in this way, give wrong data to the base stations or sinks. As sensor hubs ordinarily have short scope of transmission and rare asset, an assailant with high preparing power and bigger correspondence range could assault a few sensors in the meantime to change the genuine data amid transmission.

Sybil Attack

Much of the time, the sensors in a remote sensor system may need to cooperate to achieve an assignment, thus they can utilize appropriation of sub undertakings and excess of data. In such a circumstance, a hub can put on a show to be more than one hub utilizing the personalities of other true blue hubs (Figure 1). This kind of assault where a hub manufactures the personalities of more than one hub is the Sybil assault. Sybil assault tries to corrupt the respectability of information, security and asset use that the disseminated calculation endeavors to accomplish. Sybil assault can be performed for assaulting the dispersed stockpiling, directing component, information accumulation, voting, reasonable asset designation and bad conduct location. Essentially, any distributed system (particularly remote adhoc systems) is defenseless against sybil assault. Notwithstanding, as WSNs can have a type of base stations or entryways, this assault could be avoided utilizing productive conventions. Douceur demonstrated that, without a sensibly concentrated expert, sybil assaults are constantly conceivable aside from under outrageous and improbable presumptions of asset equality and coordination among substances. Notwithstanding, location of sybil hubs in a system isn't so natural. Newsome utilized radio asset testing to recognize the nearness of sybil node(s) in sensor arrange and demonstrated that the likelihood to distinguish the presence of a sybil hub is:

$$Pr(detection) = 1 - (1 - \sum_{all S, M, G} \frac{\binom{s}{S} \binom{m}{M} \binom{g}{G}}{\binom{n}{c}} \frac{S - (m - M)}{c})^r$$

Where, n is the quantity of hubs in a neighbor set, s is the quantity of Sybil hubs, m malignant hubs, g number of good hubs, c is the quantity of hubs that can be tried at once by a hub, of which S are Sybil hubs, M are pernicious (flawed) hubs, G are

great (revise) hubs and r is the quantity of rounds to repeat the test.

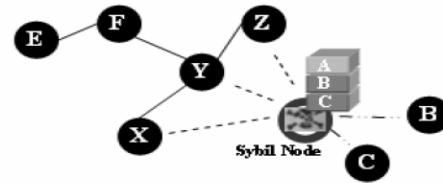


Figure 1: Sybil Attack

Black hole/Sinkhole Attack

In this assault, a pernicious hub goes about as a dark gap to draw in all the movement in the sensor organize. Particularly in a flooding based convention, the assailant tunes in to demands for courses at that point answers to the objective hubs that it contains the high caliber or briefest way to the base station. Once the vindictive gadget has possessed the capacity to embed itself between the imparting hubs (for instance, sink and sensor hub), it can do anything with the parcels going between them. Truth be told, this assault can influence even the hubs those are extensively a long way from the base stations. Figure 2 demonstrates the reasonable perspective of a black hole/sinkhole assault.

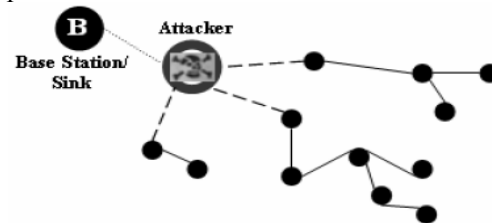


Figure 2: Conceptual view of Blackhole Attack

Hello Flood Attack

Hi Flood Attack is presented in this assault utilizes HELLO bundles as a weapon to persuade the sensors in WSN.

In this kind of assault an aggressor with a high radio transmission (named as a PC class assailant in) range and preparing power sends HELLO bundles to various sensor hubs which are scattered in a substantial territory inside a WSN. The sensors are along these lines convinced that the enemy is their neighbor. As a result, while sending the data to the base station, the casualty hubs endeavor to experience the assailant as they realize that it is their neighbor and are eventually parodied by the aggressor.

Wormhole Attack

Wormhole assault is a basic assault in which the aggressor records the parcels (or bits) at one area in the system and passages those to another area. The burrowing or retransmitting of bits should be

possible specifically. Wormhole assault is a noteworthy danger to remote sensor systems, since; this kind of assault does not require bargaining a sensor in the system rather, it could be performed even at the underlying stage when the sensors begin to find the neighboring data.

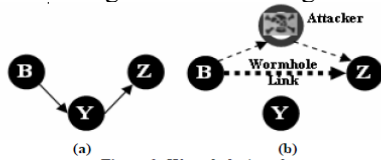


Figure 3: Wormhole Attack

Figure 3 (an and b) demonstrates a circumstance where a wormhole assault happens. At the point when a hub B (for instance, the base station or some other sensor) communicates the steering demand bundle, the assailant gets this parcel and replays it in its neighborhood. Each neighboring hub accepting this replayed bundle will view itself as in the scope of Node B, and will check this hub as its parent. Henceforth, regardless of whether the casualty hubs are multi-hop separated from B, assailant for this situation persuades them that B is just a solitary jump far from them, along these lines makes a wormhole.

In the current years, remote sensor arrange security has possessed the capacity to draw in the considerations of various analysts around the globe. In this segment we audit and guide different security plans proposed or executed so far for remote sensor systems. The below tables were taken from the source. (Sen 2010)

| Attacks on various layers of a WSN and their countermeasures | | |
|--|---|---|
| Layer | Attacks | Defense |
| Physical | Jamming | Spread-spectrum, priority messages, lower duty cycle, region mapping, mode change |
| Link | Collision Exhaustion Unfairness | Error-correcting code Rate limitation Small frames |
| Network | Spoofed routing information & Selective forwarding Sinkhole Sybil Wormhole HELLO Flood Acknowledgment flooding | Egress filtering, authentication, monitoring Redundancy probing Authentication, monitoring, redundancy Authentication, probing Authentication, packet leases by using geographic and temporal info Authentication, verify the bi-directional link authentication |
| Transport | Flooding De-synchronization | Client puzzles Authentication |

| Name | Description |
|---------------------------------------|---|
| Traffic analysis | Traffic analysis is the process of catching and investigating communication posts in order to presume information from patterns in communication [17]. |
| Denial-of-service attack (DoS attack) | It is an effort to make a computer sources unavailable to its anticipated users [18]. Builders of DoS attacks normally corrupt sites or high-profile web servers such as banks, credit card payment gateways, and domain name. |
| Replay attack | A replay attack is a violation of protection system in which relevant data is stored without approval and then present to scam the recipient into illegal procedures such as false recognition or authentication or a replicate operations [19]. |
| Interference and Jamming | Radio signals can be jammed or interfered with, which causes the message to be corrupted or lost. If the intruder has an influential transmitter, then it will be generated a strong signal to overpower the targeted signals and disturb communications [20]. These types of signal jamming are known as random noise and pulse. |
| Data forwarding phase | In the network layer, some attacks hit data packet forwarding phase. In this phase, malicious nodes do not send the data packets constantly according to the routing table. Malicious nodes simply drop data packets without any acknowledgment, change data material, hold-up forwarding real-time data packets selectively or insert garbage packets [21]. |
| Rushing attack | Two schemed attackers use the tunnel process to make a wormhole. The tunneled packets can propagate faster if a fast transmission path and dedicated channel shared by attackers, exists between the two finishes of the wormhole, rather than a normal multi-hop route. This causes the rushing attack. These attacks can act as a valuable denial of service attack beside all currently proposed on-demand WSN routing protocols [22]. |
| Resource Consumption Attack | In Resource consumption attack a compromised node can try to use battery life by forwarding needless packets to the fatality node [23]. |
| Session hijacking | In the TCP session hijacking attack, the attacker take-offs the sufferer's IP address determines the correct sequence number (expected by the target) and then performs a DoS attack on the sufferer. A session hijacking over UDP is the same as over TCP, apart from that UDP attackers do not have to worry about the transparency of managing sequence numbers because it is a connectionless protocol [24]. |
| Malicious code attacks | Malicious code (viruses, worms, spyware, and Trojan Horses) can attack both operating systems and user applications. Typically these malicious programs can spread itself through the network and cause to slow down or even damage the computer system and networks [25]. |
| Location disclosure attack | An attacker discloses information about the position of nodes or the composition of the network such as a route map and then plans further attack scenarios [26]. |

Security Schemes for Wireless Sensor Networks

Gives an investigation of secure directing in remote sensor systems. Concentrates how to configuration secure appropriated sensor systems with numerous supply voltages to lessen the vitality utilization on calculation and in this manner to expand the system's lifetime. Goes for expanding vitality productivity for enter administration in remote sensor systems and employments.

System demonstrate for its application thinks about DoS assaults against various layers of sensor convention stack displays a mapping convention which recognizes a stuck district in the sensor system and evades the broken area to keep steering inside the system, accordingly handles DoS assaults caused by sticking.

Table 1: Different Security Techniques applied to Wireless Sensor Networks

| Security Methods | Attacks | Main Features |
|------------------|------------|------------------------------|
| JAM | Dos Attack | Point to point nodes used to |

| | | |
|---|---|---|
| | | stop avoidance of the jammed region. |
| Based on wormhole | Dos Attack | Utilizes wormholes to avoid jamming |
| Random key pre-distribution, radio resource testing, etc. | Sybil Attack | By using radio resources, random key pre-distribution, registration procedure, verification of position, and code testing Sybil entity attacks are detecting. |
| Two-directional verification Multi-base station routing, multi-routing | Hello flood Attack | Two-directional verification and multiple base station routing and multi routing are used. And also adopts a secret, probabilistic, sharing compartment. |
| Based on communication security | Information or data spoofing | Efficient use of the resources. Protects the network even if part of the network is compromised. |
| Pre-distribution of random key | Data and information spoofing. Attacks information in transit | Provides flexibility in the network protects the network, even if part of the network is compromised, provides authentication measures for |

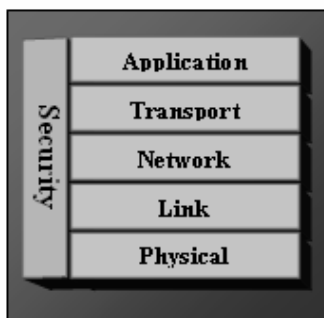
| | | |
|--------------------|---|--|
| | | senor nodes. |
| REWARD | Black-hole attacks | Uses geographic routing and takes advantage of being the sender to see the nearer transmission and detects black-hole attacks. |
| TinySec | Data and information spoofing, the messages repeat the attacks. | Centered on providing message authenticity, integrity and confidentiality messages works in the link layer. |
| SNEP y μ TESLA | Data and information spoofing, the messages repeat the attacks. | Semantic security, Replay protection, data authentication, low communication overhead. |

REWARD is a steering calculation which battles against dark gaps in the system. Proposes isolate security plans for information with different affectability levels and an area based plan for remote sensor arranges that ensures whatever remains of the system, notwithstanding when parts of the system are traded off. Executes symmetric key cryptographic calculations with deferred key exposure on bits to build up secure correspondence channels between a base station and sensors inside its range. What's more, propose key pre-appropriation plans, which focus to enhance the flexibility of the system. In Table 1 we condense different security conspires alongside their fundamental properties proposed so far for remote sensor systems.(Kaschel, Mardones et al. 2013)

All-encompassing Security in Wireless Sensor Networks

An all-encompassing methodology goes for enhancing the execution of remote sensor systems as for security, life span and network under

changing natural conditions. The comprehensive approach of security worries about including every one of the layers for guaranteeing general security in a system. For such a system, a solitary security answer for a solitary layer won't not be a productive arrangement rather utilizing a comprehensive approach could be the best choice.



The comprehensive approach has some essential standards like, in a given system; security is to be guaranteed for every one of the layers of the convention stack, the cost for guaranteeing security ought not outperform the evaluated security chance at a particular time, if there is no physical security guaranteed for the sensors, the safety efforts must have the capacity to show an elegant debasement if a portion of the sensors in the system are traded off, out of request or caught by the foe and the safety efforts ought to be produced to work in a decentralized manner. On the off chance that security isn't considered for the majority of the security layers, for instance; if a sensor is some way or another caught or stuck in the physical layer, the security for the general system breaks in spite of the way that, there are some effective security components working in different layers. By building security layers as in the comprehensive approach, assurance could be built up for the general system.

Conclusion

The vast majority of the assaults against security in remote sensor systems are caused by the addition of false data by the traded off hubs inside the system. For shielding the consideration of false reports by bargained hubs, a method is required for distinguishing false reports. In any case, growing such a discovery component and making it effective speaks to an extraordinary research challenge. Once more, guaranteeing comprehensive security in remote sensor arrange is a noteworthy research issue. A considerable lot of

the present proposed security plans depend on particular system models. As there is an absence of joined push to take a typical model to guarantee security for each layer, in future however the security components turn out to be entrenched for every individual layer, consolidating every one of the systems together to make them work, as a team with each other will cause a hard research challenge. Regardless of whether all-encompassing security could be guaranteed for remote sensor arranges, the cost-adequacy and vitality productivity to utilize such systems could in any case posture incredible research challenge in the coming days.

References

- Abd-El-Barr, M. L., Al-Otaibi, M. M., & Youssef, M. a. (2005). Wireless sensor networks- part II: routing protocols and security issues. *Canadian Conference on Electrical and Computer Engineering, 2005.*, (May). <https://doi.org/10.1109/CCECE.2005.1556879>
- Anjali, Shikha, & Sharma, M. (2014). Wireless Sensor Networks : Routing Protocols and Security Issues. *5th ICCNT*, 3–7. <https://doi.org/10.1109/ICCCNT.2014.6962992>
- Bin Lu, Habetler, T. G., Harley, R. G., & Gutierrez, J. A. (2005). Applying Wireless Sensor Networks in Industrial Plant Energy Management Systems - Part I: A Closed-Loop Scheme. *IEEE Sensors, 2005.*, 2005, 145–150. <https://doi.org/10.1109/ICSENS.2005.1597657>
- Dr. G. Padmavathi, M. D. (2009). A Survey of Attacks, Security Mechanisms and Challenges in Wireless Sensor Networks. *International Journal of Computer Science and Information Security, IJCSIS, Vol. 4, No. 1 & 2, August 2009, USA*, 9.

KASCHEL, H., MARDONES, J. and QUEZADA, G., 2013. *Safety in Wireless Sensor Networks: Types of Attacks and Solutions. Studies in Informatics and Control*, 22(3), pp. 323-329.

SEN, J., 2010. *A survey on wireless sensor network security. arXiv preprint arXiv:1011.1529*.

An Exposition of Performance Comparison of Graphic Processing Unit Virtualization Techniques

¹Asif Farooq, ²Tahir Iqbal

¹Lecturer, Department of Computer Science and Information Technology,
University of Sargodha, Lahore Campus, Pakistan

²Sr. Assistant Professor, Department of Computer Science and Information Technology,
Bahria University, Lahore Campus, Pakistan

E-mail: ¹asiffarooq@uoslahore.edu.pk, ²Tahir.iqbal@bahria.edu.pk

Abstract— As the demand for computing power is increasing the number of new and improved methodologies in computer architectures are expanding. With the introduction of accelerated heterogeneous computing model, compute times for complex algorithms and tasks are reduced significantly as a result of high degree data parallelism. GPU based heterogeneous computing can not only benefit Cloud infrastructures but also large-scale distributed computing models to work more cost-effective by improving resource efficiencies and decreasing energy consumptions. Thus to implement such paradigm on cloud and largescale infrastructure would require effective GPU virtualization techniques. In this survey, an overview of GPGPU virtualization techniques using CUDA programming model is reviewed with a detailed performance comparison.

Keywords: GPU, HPC, Virtualization, Cloud, GPGPU, CUDA

I. INTRODUCTION

High-Performance Computing (HPC) model is adopted by combining two different architectures: one is multicore processors with high-performance general-purpose cores and second is many-core accelerators which are generally high-performance GPUs. Hence combining multi-core CPU with GPU gives immense data parallelism through a large number of simple accelerators for intensive tasks and applications. This combination of CPU and GPU is so effective that a lot number of supercomputers have adopted this approach to increase performance. According to a report on Top500 [1, 2], both Google and Microsoft have added high-performance GPU to their Cloud infrastructures to increase computation performance.

For such environments to provide virtualization services, GPU virtualization plays an important role. Mainly there are two types of GPU virtualization methods: GPU pass through and GPU sharing. In GPU pass through a single virtual machine (VM) can access a GPU directly while in GPU sharing refers when a single GPU is shared by multiple VMs. Implementation solutions like Virtual Dedicated Graphics Acceleration (vDGA) and Virtual Shared Graphics Acceleration (vSGA) are provided for GPU pass-through and GPU sharing respectively by VMware Inc. Implementation techniques are further classified into three manners:

- API remoting
- Para and full virtualization
- Hardware-supported virtualization

- Hardware-supported virtualization

API remoting is the most common implementation technique adopted by many authors. In this methods guest, OS GPU calls are forwarded to host OS with the help of a wrapper library. Host OS is equipped with physical GPU(s) to intercept and execute GPU calls for guest OS.

The second approach is based on two terminologies, both of these offer GPU virtualization at the driver level, for which a custom driver is required for each specific GPU that is used in host OS. The major difference between para and full virtualization is that former a slight driver modification is required in guest OS while in latter no such modification is required.

In hardware supported virtualization, either GPU vendors or motherboard chipset provides hardware extension features which allow a guest OS to directly access a remote GPU.

All GPU virtualization techniques require some sort of programming model for the parallel processing of applications or to perform general Purpose computations on GPUs. For this Efficient and freeware programming models like CUDA and OPENCL are available that provides frameworks for applications to be executed across heterogeneous clusters.

II. RELEVANT LITERATURE

With the introduction of CUDA [3] and OpenCL [4] more and more GPGPU approaching are developing to enhance the performance of GPU virtualization. CUDA and OpenCL are the basic programming models of GPGPU in cloud computing but these two approaches are a lot less flexible than today's cloud computing requirements. Hence to improve that in a study [5] a remote CUDA approach is introduced which enables transparent and remote GPGPU acceleration in HPC clusters. The main benefit of rCUDA is that users can access any Accelerator (CUDA Compatible) from any node in the cluster. Hence rCUDA directs the CUDA API call to the remote server with GPU and perform required tasks.

Other approaches were also introduced with the aim to increase GPU virtualization efficiency and to increase the performance of HPC clusters. In a study [6] authors combine OpenACC with rCUDA framework. OpenACC is application programming interface (API) integrated with HPC architecture to allow developers to accelerate application executions.

GPGPU in HPC cluster can be maintained by any virtual machine manager, for that, a number of hypervisors are available to perform desire tasks. In a study [7] authors used an ESX Hypervisor for GPU virtualization. As explained by [7] ESX Hypervisor does not allow the multiplexing of GPUs

to be shared among VMs. A new approach vmCUDA is proposed that is flexible enough to allow CUDA application running concurrently in a number of VMs to share GPU(s).

In another study [8] an improved version of rCUDA is proposed with some significant modifications to expand the single node capabilities of original CUDA. Thus a single application running in one node can now access GPU across all nodes with low overhead.

In [9] authors combine gVirtuS (a general virtualization system introduced in 2010 mainly for x86 and ARM machines) with CUDA 6.5. Enabling GPU sharing in x86 and ARM machine.

As CUDA and rCUDA and other similar virtualization approaches need some sort of communication between guest OS and host OS to perform GPGPU. Most of these approaches used TCP/IP for communications, which might cause higher overhead values. To resolve such issue authors in [10] used InfiniBand Connect-IB network adapters (compatible with CUDA and rCUDA) and show the effectiveness in overhead values by experimentations.

Besides using CUDA as a basis many authors also proposed Improved Virtualization techniques using some of which increase performance by improving the GPU resource scheduling process GPU scheduling also have a great impact on the performance of GPU virtualization. Many GPU scheduling approaches have been proposed lately as the GPGPU becomes more common. In such a study [11] proposes a vGASA approach for GPU scheduling with feedback control feature. Experiments show the overhead is greatly reduced when implemented in cloud gaming scenario. In another similar study authors [12] proposed a VGRIS (Virtualized Graphics Resource Isolation and scheduling). This approach mainly depends on para Virtualization architecture.

III. TECHNICAL REVIEW

Since the advent of CUDA more and more researchers seemed to be focused on GPGPUs to accelerate compute-intensive applications. Following is a list of handful techniques for improving GPGPU performances and efficiency:

GVIM

In [13] authors implemented CUDA API using Xen hypervisor in their proposed technique GVIM. The main focus of GVIM was to efficiently manage resources between virtual machines during general-purpose GPU computations. As proposed in 2009 GVIM uses CUDA 1.1 library to provide GPU access to guest virtual machines. Experiments show improvements in overhead and computation time in both virtualized and non-virtualized systems.

vCUDA

In [15] authors focuses on giving hardware acceleration control to applications within virtual machines to increase the overall computation performance for applications. VCUDA consists of three modules:

- vCUDA library
- virtual GPU
- vCUDA stub

As CUDA framework is flexible enough for researchers to allow custom libraries to be used instead of standard CUDA provided libraries. VCUDA utilizes this benefit and replaced standard CUDA library in the guest operating system to handle and redirects all the GPU calls to the core vCUDA stud, which not only keep the log of each activity but also monitors each application GPU calls and performs the primary task of sending back the results to guest OS. The evaluation showed performance improvements in VMs as compared to non-virtualized environments.

gVirtuS

In [16] authors try to reduce the gap between in-house HPC clusters and commercially available virtual clusters by proposing techniques called GPU Virtualization Service (gVirtuS). This approach allows transparent access to a VM without using hypervisor.

rCUDA

As each node in HPC is equipped with multiple GPUs to accelerate application. The major disadvantage of which is power consumption. According to a report [18], NVidia GTX 1080 would consume up to 150 watts which is a lot more than a typical CPU. Thus to reduce power consumption we need to reduce the number of GPUs in each cluster without reducing the overall system performance. To resolve the major issue authors in [17] presented such an approach called remote CUDA (rCUDA). The main idea behind the approach was to make each GPU accessible to every node, for this proposed framework to work an array of virtual CUDA compatible GPUs are placed in those machines that do not have a physical GPU installed.

DS-CUDA

In another similar study [19] a distributed middleware was proposed which virtualizes GPUs in a cloud environment is such a way that GPU appears to be installed on local machines. This approach reduces the overall cost of the system and induced great reliability for consumer GPUs.

LoGV

In this paper [20] authors presented a low overhead GPU virtualization (LoGV) technique. This approach uses host drivers and CUDA runtime without modification and enables GPGPUs sharing for VMs. LoGV relies on the hypervisor to grant resources secured by itself to VMs for mutual protections between hardware. LoGV consists of two major parts:

- Guest Kernel Driver
- Extension to Hypervisor

The guest kernel driver performs operations directly linked to VMs while the hypervisor extension manages and allocated resources for application protections

Grid CUDA

Authors in [21] proposed programming toolkit for programmers with a primary goal to provide a platform for its user to write a program with CUDA API and execute these programs using GPGPU. Grid CUDA also supports parallel

executions by distributing programs over multiple GPUs. The author evaluated their proposed approach by integrating it into Teamster-G (real Computational Grid).

IV. COMPARATIVE ANALYSIS

Multiple parameters are selected to compare above-mentioned GPU virtualization approaches. As all the approaches mentioned in the technical review are CUDA compatible hence only GPGPU parameters are compared. To further compare the performance of each of above NVidia benchmarks results from selected approaches are presented as below:

VERSION COMPATIBILITY:

Different virtualization frameworks offer completely different options, for example, the vCUDA technology supports the recent CUDA 3.2 version capable of running with any CUDA runtime API libraries. As compared to rCUDA its communication shows slightly larger overhead values than CUDA. GVim works with CUDA1.1 which was a lot less flexible than the more recent version and doesn't implement the complete runtime API. The gVirtuS approach uses a more recent version than GVim as it is compatible with CUDA 2.3. This approach is implemented using only a small portion of

HYPERVISOR LIMITATION:

Performance results of overall approaches greatly depend on the type of hypervisor used during the performance benchmarks of CUDA compatible approaches. Thus the selection of hypervisor also have a great impact on the CUDA GPGPU performance, as states by multiple authors [8,15,22] in their research that the overall performance of each proposed technique mainly depends on the following :

- CUDA API version
- Communication controller
- Virtualization manager (Hypervisor)

As reviewed from articles the most commonly used hypervisor are:

- KVM
- VMWare
- XEN
- Linux Containers

VMWare's hypervisors provide a steady performance when tested under multiples virtualization variation. As multiple approaches [9, 19, 22] used VMWare as a hypervisor for CUDA GPGPU benchmarking, results from these authors shows a balanced performance and acceptable overhead values. XEM hypervisor also provides the same performance as VMWare. The major difference between two was the compatibility with CUDA API and support for GPU Pass-through as both of this hypervisor nearly correlates each other in result regarding overhead values as tested by [22]. KVM hypervisors work well if used lower version of CUDA with TCP/IP protocol for communication. More advanced controllers are not supported like INFINIBAND. Studies [19,

CUDA runtime libraries. vGPU uses a more advanced version of CUDA. It utilizes the flexibility of CUDA 4.0 but the performance data provides by authors in this approach is a little fuzzy hence not included in this comparison review. GridCUDA supports the recent CUDA 2.3 as compared to DS-CUDA, which integrates a more advanced and improved version of CUDA (4.1) libraries. One of the main advantages of CUDA more recent version was that includes support for State of the Art network controllers called INFINIBAND, which provides best data rate transfer with very little overhead.

Table 1: VERSION COMPARISON

| Technique | Hypervisor | CUDA version | Open Source | Remote acceleration |
|-----------|-------------|--------------|-------------|---------------------|
| GVim | Xen | 1.1 | | |
| vCUDA | Xen | 3.2 | | |
| gVirtuS | VMware, Xen | 2.3 | ✓ | ✓ |
| rCUDA | KVM | 6.5 | | ✓ |
| DS-CUDA | VMware | 4.1 | ✓ | ✓ |
| LoGV | KVM | - | | |
| Grid CUDA | - | 2.3 | | |

the price of flexibility and security, however. LXC was a less Versatile than most full virtualization hypervisors

PERFORMANCE COMPARISON:

To measure each reviewed approaches performance following two factors are mainly used

- Overhead
- Execution Time (ms)

Following table shows an overall performance comparison.

Table 2: PERFORMANCE COMPARISON

| Technique | Overhead SLA percentage | Execution (ms) |
|-----------|-------------------------|----------------|
| GVim | 67 | 4.5 |
| vCUDA | 35 | 3.8 |
| gVirtuS | 48 | 8.0 |
| rCUDA | 18 | 2.2 |
| DS-CUDA | 20 | 2.5 |
| LoGV | 76 | 6.8 |
| Grid CUDA | 35 | 2.6 |

Table 2 shoes Overhead SLA violation values as well as execution time. Higher SLA value means low bandwidth thus the lower the percentage the better the results of bandwidth. Similarly, the execution time is also reviewed from studies. Computational cost is calculated along differ parameters in

GPGPU virtualization. The lower the execution time the better the results. Following chart shows an overview of comparison.



Figure 1 Performance Comparison

V. CONCLUSION:

As more and more researchers are focused on making high-performance computing more efficient in terms of performance and energy, new and improved approaches are emerging to accelerated compute intensive application. In that case GPGPUs are the primary focus for HPC. In the survey multiple CUDA related approaches are reviewed and compared side by side along with multiple parameters related to GPGPUs. Moreover a performance comparison is also listed that overviews two most important parameters in GPGPU computing: overhead and execution time. All the results are shown elaborately to summarize the comparison.

REFERENCES:

- [1]: <https://www.top500.org/news/google-rolls-out-gpu-cloud-service/>.
- [2]: <https://www.top500.org/news/microsoft-adds-high-performance-gpus-to-azure-cloud/>.
- [3]: NVIDIA Corporation, CUDA API Reference Manual Version 5.5 2013.
- [4]: A. Munshi (Ed.), The OpenCL Specification Version 1.2, Khronos OpenCL Working Group, 2011.
- [5]: A complete and efficient CUDA-sharing solution for HPC Clusters, Antonio J. Peña
- [6]: Exploring the Suitability of Remote GPGPU Virtualization for the OpenACC Programming Model Using rCUDA, Adrian Castell
- [7]: GPU Virtualization for High-Performance General Purpose Computing on the ESX Hypervisor, Lan Vu
- [8]: Improving the user experience of the rCUDA remote GPU virtualization framework, Carlos Reaño.
- [9]: On the Virtualization of CUDA Based GPU Remoting on ARM and X86 Machines in the GVirtuS Framework Raffaele Montella
- [10]: POSTER: Boosting the Performance of Remote GPU Virtualization Using InfiniBand Connect-IB and PCIe 3.0, C. Reano.

- [11]: vGASA: Adaptive Scheduling Algorithm of Virtualized GPU Resource in Cloud Gaming, Chao Zhang.
- [12]: VGRIS: Virtualized GPU Resource Isolation and Scheduling in Cloud Gaming, Miao Yu
- [13]: GViM: GPU-accelerated Virtual Machines, Vishakha Gupta
- [14]: <https://developer.nvidia.com/cuda-zone>.
- [15]: vCUDA: GPU-Accelerated High-Performance Computing in Virtual Machines, Lin Shi
- [16]: A GPGPU Transparent Virtualization Component for High-Performance Computing Clouds, Giulio Giunta
- [17]: rCUDA: Reducing the Number of GPU-Based Accelerators in High-Performance Clusters, Jose Duato
- [18]: <https://www.computerpoweruser.com/article/23395/power-supply-requirements-for-new-nvidia-and-amd-gpus>.
- [19]: DS-CUDA: a Middleware to Use Many GPUs in the Cloud Environment, Minoru Oikawa
- [20]: LoGV: Low-overhead GPGPU Virtualization, Mathias Gottschlag
- [21]: GridCuda: A Grid-enabled CUDA Programming Toolkit, Tyng-Yeu Liang
- [22]: GPU Passthrough Performance: A Comparison of KVM, Xen, VMWare ESXi, and LXC for CUDA and OpenCL Applications, John Paul Walters.

User Feedback based Ranking for Yahoo Answers

Muhammad Awais Shaheen, Student
Department of Computer Sciences
International Islamic University
Islamabad, Pakistan
awais.mscs823@iiu.edu.pk

Abstract- Web sites, such as, Yahoo Answers, encyclopedias, blogs and web forums which are described by a loose edit control, which permits anybody to freely alter their content. As an outcome the quality of this content raises much concern. To control content quality issue numerous of manual (best answer, up voting) quality control mechanisms introduced. In any case, given their size and change rate, manual evaluation systems don't scale the content that is new and time to time assessed. This has a negative effect on the ranking and recommendation system. So, we used automatically ranking technique to rank the yahoo answers. We use support vector regression method, used text feature, main idea is to rank new answers or edited answers with the help of their text. In yahoo answers, after manual rank the answer on the top, new answer does not rank on top even it has good content. This problem is facing by many sites like Yahoo Answers. This is a biased toward best answers selection and got up-vote. We used NDCG (normalized discounted cumulative gain) method to compare results because it measures the gain or effectiveness of answers at its existing position.

Keywords- Ranking; Content Quality Assessment; Answer Quality

I. INTRODUCTION

Our contribution in this paper, we rank the new answers of Yahoo Q&S, use automatically ranking method (support vector regression) rather than manual method, using two (Random forest and Support Vector Regression) different method on Yahoo dataset to evaluate the better result, also weight (using information gain and information gain ratio) the feature to use in ranking process so that we will use these feature that help us to get optimize result. We also apply methodology individual and separate set of features, which explain the features importance. Bias is defined as "Cause to feel or show inclination or prejudice for or against someone or something" [7]. Biases often effect the rank of answers, for example if an answer is already manually select the top rank then novel answers cannot rank better against already selected top rank answer. There are many kinds of biases and effect the result accuracy. Many of result are taken by searcher are wrong. But searcher is satisfied by result because of their own belief and biases [1]. Biases which are discussed in our research are related to rank the document, it does not relate to culture, religious. The parameter used to check biases are related to ranking the document.

We explore the yahoo answer and find many biases like asker's satisfaction and up-voting. In asker's satisfaction, question asker can select the answer is best, it is basically showing that asker satisfy with this answer or content of answer. In up-voting, different user can vote against answers, it shows that people are satisfied with this answer. In both above case novel or late comer answer effect bias move above mention case [12]. Satisfaction is a feedback of user/rater. Everyone has its own thinking and he replied against question that is best he thinks or selects the best or up-vote. Its mean this person is satisfied with this answer/replied. But asker satisfaction is considered to be best answer selected [8]. Yahoo! Answers formerly known as Yahoo Answers is a community driven question and answer site or a knowledge market from Yahoo! That allows users to both submit questions to be answered and answer questions asked by other users and search the already existed question and satisfy you [8]. The working of yahoo is very simple. People search

already exist question if exist they go with them otherwise they ask question and wait for answer.

Table 1. Yahoo Levels

| Level | Points | Questions | Answers | Follows |
|-------|-------------|-----------|---------|---------|
| 7 | 25000 | 20 | 160 | 100 |
| 6 | 10000-24999 | 20 | 160 | 100 |
| 5 | 5000-9999 | 20 | 160 | 100 |
| 4 | 2500-4999 | 20 | 160 | 100 |
| 3 | 1000-2499 | 15 | 120 | 100 |
| 2 | 250-999 | 10 | 80 | 100 |
| 1 | 1-249 | 5 | 20 | 100 |

Level is another way to keep track of how active have been. The most point you accumulate, the higher your level. Yahoo answers recognize your level achievements with our special brand. Finally, as you attain higher level, you will be able to contribute more to Yahoo answers. You can ask and rate more frequently. This is the yahoo answer level system how people answer different question and reach at top contributor [8].

Table 2. Yahoo Answers Point

| Action | Points |
|--|----------------|
| Being participant on yahoo answers | One time 100 |
| Ask a question | -5 |
| Choose a best answer for your question | 3 |
| Answer a question | 2 |
| Self-deleting an answer | -2 |
| Login to yahoo answer | Once daily |
| Have your answer as best | 10 |
| Receive a thumb up | 1 per thumb up |
| Receive vote | -10 |

This is yahoo point system. How point given to people and reach top contributor [8]. Did you receive an answer to your question that was right on target, let the answerer and the answers community know it was helpful y choosing it as “BEST ANSWER”. You have option to choose a “BEST ANSWER” one hour later after asking your question.

- a. Find and select My Question
- b. Click the question
- c. Find the answer you like best
- d. Click choose as best answer
- e. Rate the best answer
- f. Click submit

Success! The Best answer will move on the top of the answer module.

In yahoo answer there are many existing question. When we query to search answer it shows result on the basis of relevance or older there is manual ranking system [2]. In case of ask question, People give answer and some people who are satisfied with them like the answer and finally asker select best answer and rate it.

The problem is to rank answer of Yahoo Answers where user has accessed text/answer freely edited or add answer against questions. With the passage of time more answer is added or edit against question, it does not mean all text is good; we have automatic method to rank the content instead of manual. Because through manual method it is difficult to measure the quality of updated or newly added content. Through manual method many biases are added in content [22].

Learn to rank technique is used to rank the content; point wise approach is to because the values consist of numeric data. Support vector regression method is used with help of R to rank the answer. We used text feature for this with Support Vector Regression (SVR). Attribute consist of numeric values that is why used regression. In previous methodology (Random Forest), stack flow dataset is rank with text feature and give good result. In our methodology, we used dataset (yahoo answers) and method that is SVR to rank. We also compare with previous methodology [23]. Normalized discounted cumulative gain (NDCG) method is a measure of ranking quality. It measures the gain of answers at their present position [25]. In previous method it shows good gain result at top k. the result is 21% better with other method results [22].

2. RELATED WORK

In this section, I read many paper and in the base paper Ryan W. White said Biases can be observed in information retrieval (IR) in situations where searchers seek or are presented with information that significantly deviates from true likelihoods [2]. There are many biases present in this rate Paper such as belief in biases, confirmation on contradiction sentences are the main biases explain in this paper. In [3] Ryan W. White talks over the pre-belief [3]. Earlier work demonstrates that affirmation inclination, characterized as the propensity to look for affirming proof, is predominant on the Web also. While this has been attributed to the person's mental needs or psychological constraints, the parts of web indexes and seek settings have generally been disregarded. The objectives of this examination are to analyze how look settings may change the creation of indexed lists, and how, if at all web search engines may add to affirmation inclination [4]Biases in yahoo like Positional bias, Confirmation, Crowd Source Judgment, Best Answer, Physician Answer. Positional bias, there are more chance of top document are selected. In daily life we see that top document are more selected and if there is bias in it then it leads toward negative information [1]. Confirmation bias, people have belief and they go to confirm their belief are select top document and further go with it and just confirm it by help of other document [1]. Crowd source judgment, answers given by different people and then people like and unlike it. It creates confusion to select the best answer. Basically best answer is also a bias for satisfy answer. It is often attractive thing [20]. Physician answer, physician answer is a good way to answer a question but often they answer with their own satisfaction. If two or more satisfy with same answer then possibility of correctness will be increase [20]. In Yahoo answer for sentence retrieval presented the idea to select the most relevance answer against the question. They used idea of sentences retrieval rather than words match. They used combination of two method class based method and train trigger method and gain accuracy of 19.29% than individually used [2]. In belief and dynamic biases, multiple additive regression tree methodology to classification of data and root mean square error to evaluate accuracy. Main theme of paper is to classify data between biased and un-biased data using user behavior on the web page [16]. Crowd source internet poll presented the idea to evaluate of user satisfaction or success on internet. They used average user satisfaction and Success Index method to rate the post [17]. In toward less biased we search, xiting liu present idea that biased information lead toward unwise decision or undesirable consequences. To achieve less biased result use novel text extraction and aggregative prediction [18].

There are many kind of biases find in search existing answer question. Like search result on basis of relevance [1]. The result shown own query of search question is basis of relevance or older time question. There is no ranking system of satisfaction to show that which page is satisfy more person. Best answer, it is basically selected by asker and creates biases for search question [20]. Actually it is selected by its own beliefs. Even we see biases like top contributor answer. Asker gives more attention to top contributor answer. Conformity is also issue of people search for conform the answer, not go to contradiction statement [1]. People often move toward positive information and neglect negative information.

Biases in answer of asker question. Top contributor answer biases, asker often look toward top contributor [20]. Biases in crowd source judgment, attract toward question that is like by many people. Unlike option is also issue, for example one question have 20 like and second has 10 like, if the person like second question also unlike first question [20]. It creates negative impact on first often in this situation asker go to second question.

III. PROBLEM STATEMENT

A. Definition 1

Kimcapps explains the yahoo rank the answers by selecting the BEST answer. Let suppose $a_i = \{a_1, a_2, a_3, a_4, a_5\}$ are the answers against a question and b is the best answer. If asker selects $a_i = b$, it supposed to a best answer and placed at the top. Actually the answer is select manually and later feedback after selects the best answer never rank as best even it contain better content [20].

B. Definition 2

In paper “Measuring explicit user feedback” Apostolos explains rank answers by the explicit evolution measure by using Average User Satisfaction method. Suppose $a_i = \{a_1, a_2, a_3, a_4, a_5\}$ are answers against a question, $uv_i = \{8, 5, 6, 1, 0\}$ are the vote against the answer. Votes of user require ranking the post $\frac{\sum_{visted\ post\ vote}}{|visted\ post|}$, in this process it is obvious older answer have maximum vote than new answer [17].

C. Definition 3

In paper [30], Martha Sideri use implicit method rank answers $a_i = \{a_1, a_2, a_3, a_4, a_5\}$ against question q , he used the Success Index method to rank the answer it is also click based, suppose there are $n=10$ post and user click $n=2$ and 10. If $n=2$ click first so SI is 27.5%, if second then 17.5%. Actually it rank depend on the position and user behavior. It is also manually work, before this SI is zero for all post [17].

D. Statement

To rank the yahoo answers against a question, so user can facilitate with answer at the top that is rich with content. The main problem is to rank novel answers in the yahoo Q&S, rank newly added answers, updated answers, novel content or answers provided after selected the best answer, deal with manually ranked answers where sites are characterized with lose edit control. The problem statement can be confined to two research questions:

- a. How to rank answers where manually select the BEST answer.
- b. How to rank the new answer where answers are based on user's behavior (manually rank).

IV. METHODOLOGY

A. Baseline 1(Random Forest)

Decision tree learning uses the decision tree to conclude the target value. It is a predictive approach used in machine learning. There are different type of value used in tree model (discrete and continue). Tree model uses classification for discrete value and regression for continue value. Classification predicts the data belong to a specific class and regression predicts data a real number [24]. Random forest is ensemble learning method, built multiple decision trees by continually resampling training data with replacement and voting the trees for a consensus prediction. It is also known as bootstrap aggregated or bagged. It is used for both classification and regression [24].

Random forest first introduced by Tin Kam ho using random subspace method and extension is added to Breiman and Adele Culter and random forest is trademark; also add the idea of bagging and random selection of feature. Later Amit and Geman construct collection of tree and controlled variance [24].

Input: $D = \{(x_1, r_1), \dots, (x_n, r_n)\}$, $K : 0 < K \leq f, M > 0$

Output: $T(\cdot)$

```

1: for  $i = 1$  to  $M$  do
2:    $D_t \leftarrow \text{sample}(D)$ 
3:    $h_t \leftarrow \text{BUILT\_DECISION\_TREE}(D_t, k)$ 
4: end for
5:  $T(\cdot) \leftarrow \frac{1}{M} \sum_{t=1}^M h_t(\cdot)$ 
6: return  $T(\cdot)$ 

```

Let discuss the algorithm of Random Forest, $D = \{(x_1, r_1), \dots, (x_n, r_n)\}$ are the subset of document and their relevant rank. D_t is the sample of document that is selected, h_t is the decision tree that is made by selected document and elected feature, K are the feature, M represent the subset. The overall summary is we take subset of document $D = \{(x_1, r_1), \dots, (x_n, r_n)\}$ that have M subset. So we start select subset of document and features K and make the decision tree (h_t). After completion, we take average of decision tree to make prediction for test data.

B. Baseline 2 (Multiple Additive Regression Tree)

Multiple Additive Regression Trees (MART) is utilized to tackle the forecast issues in view of substantial dataset. Friedman depicts in detail the system behind this approach which broadens and enhances the CART procedure and has more noteworthy precision. It is simple actualize, programmed and keep up numerous alluring component like robustness. MART watches out for impervious to changes of indicators and reaction factors, outliers, missing values and to the incorporation of conceivably expansive quantities of immaterial indicator factors that have almost no impact on the reaction. These two last properties are specifically compelling since they are two of the best troubles when utilizing exchange information to foresee misrepresentation. In this section a brisk diagram of MART is given with specific thoughtfulness regarding translating the outcomes, deciding the impact of indicator factors on those outcomes, and estimating the significance of those factors [16].

MART is one of a class of strategies frequently alluded to as boosting. Boosting is a general technique that endeavors to boost the exactness of any given learning calculation by fitting a progression of models each having a poor error rate and afterward consolidating them to give an outfit that may perform extremely well. MART is a speculation of the tree boosting that endeavors to increment prescient precision with just a direct forfeit of the attractive properties of trees, for example, speed and interpretability. Because of the boosting procedure, MART creates a precise and successful off-the-rack technique for information mining [16].

Input: training set $\{(x_i, y_i)\}_{i=1}^n$, differentiate loss function $L(y, F(x))$, iterations M .

```

1:  $F_0(x) = \arg \min_Y \sum_{i=1}^n L(y_i, Y)$ 
2: for  $m = 1$  to  $M$ 
3:   for  $i = 1$  to  $n$ 
4:      $r_{im} = - \left[ \frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$ 
5:   End for
6:    $h_m(x) = \{(x_i, r_{im})\}_{i=1}^n$ 
7:    $Y_m = \arg \min_Y \sum_{i=1}^n L(y_i, F_{m-1}(x_i) + Y h_m(x_i))$ 
8:    $F_m(x) = F_{m-1}(x) + Y_m h_m(x)$ 
9: End for
10: Return  $F_M(x)$ 

```

We take input training set $\{(x_i, y_i)\}_{i=1}^n$, differentiate loss function $L(y, F(x))$, number of iterations M . In $F_0(x) = \arg \min_Y \sum_{i=1}^n L(y_i, Y)$ we are take initial constant value that is a point where $F_0(x)$ is minimum. $r_{im} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$ describes that we are taking pseudo residual, $h_m(x) = \{(x_i, r_{im})\}_{i=1}^n$ is used to fit a base learner to pseudo residual. $Y_m = \arg \min_Y \sum_{i=1}^n L(y_i, F_{m-1}(x_i) + Y h_m(x_i))$ solving the one dimensional problem, $F_m(x) = F_{m-1}(x) + Y_m h_m(x)$ function is used to update the model, hence return output $F_M(x)$.

C. Proposed (Support Vector Regression)

Support vector machine utilized for characterization as well as utilized for relapse. Still it contains all the fundamental highlights that describe most extreme edge calculation; a non-direct capacity is bolstered by straight learning machine drawing into high dimensional piece prompted include space. The volume is controlled by parameters that don't rely upon the dimensionality of highlight space [23].

Support Vector Machines are extremely unequivocal class of calculations, described by use of pieces, no nattendance of neighborhood minima, scantiness of the arrangement and limit control acquired by following up on the edge, or on number of help vectors, and so forth.

They were developed by Vladimir Vapnik and his collaborators, and first presented at the Computational Learning Theory (COLT) 1992 gathering with the paper. All these decent highlights however were at that point introduce in machine learning since 1960, extensive edge hyper planes use of parts and geometrical translation of portions as inward items in an element space. Comparable advancement strategies were utilized as a part of example thankfulness and meager condition systems were generally talked about. Use of loose factors to beat clamor in the information and non-detachability was likewise presented in 1960. In any case it was not until the point that 1992 that every one of these highlights were assembled to shape the maximal edge classifier, the essential Support Vector Machine, and not until 1995 that the delicate edge form was presented [23].

Input: training set $\{(x_i, y_i)\}_{i=1}^n$

Output: $f(x)r$

- 1: For $i=1$ to m
- 2: $S = \begin{pmatrix} a_i \\ c_i \end{pmatrix}$
- 3: $S^* = \begin{pmatrix} a_i \\ c_i \\ b_i \end{pmatrix}$
- 4: $w = \sum_{i=1}^m \alpha_i, S^*_i$
- 5: $w^* = \frac{1}{2} \|w\| \quad \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon \end{cases}$
- 6: $f(x) = w^* + C \sum_{i=1}^m (\xi_i, \xi^*_i) \quad \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi^* \\ \xi, \xi^* \geq 0 \end{cases}$
- 7: If(SVR==Linear)
- 8: $f(x)r = \sum_i^m (\alpha_i, \alpha^*_i) \cdot \langle x_i, x \rangle + b$
- 9: Else
- 10: $f(x)r = \sum_i^m (\alpha_i, \alpha^*_i) \cdot K \langle x_i, x \rangle + b$
- 11: End If
- 12: End for
- 13: Return $f(x)r$

In the above algorithm, we take input x_i and y_i and output should be $f(x)r$. First we talk the support vector S_i from the point, because we have find the point that show the boundary between them. There are sometimes exist bias value so we

add bias b value and take modify support vector \hat{S}_i . To find the value of \hat{w} , we add parameter α_i . ε is the distance between hyper plane and boundary, to reduce the model complexity $\|w\|$. ξ is a slack variable, to measure the training sample outside the ε zone. If data is not linearly separated then use $(x)r = \sum_i^m (\alpha_i, \alpha'_i) \cdot K(x_i, x) + b$ otherwise $(x)r = \sum_i^m (\alpha_i, \alpha'_i) \cdot \langle x_i, x \rangle + b$. The two and three dimension is computationally expensive so use kernel trick. A function takes input as a vector as an original shape and return dot product of feature base is called kernel function.

In a kernel function, take dot product between vectors set as every point is mapped into a high dimension space by some transformation. So essentially we transfer a nonlinear base to linear base.

V. Experimental Setup

A. Dataset

A data set is a collection of related, discrete items of related data that may be accessed individually or in combination or managed as a whole entity. We have taken yahoo dataset from the yahoo web site [9]. Dataset consist of text, we extract these feature from the text using formula that is commonly used and available. The extracted features are already used in a paper [22]. After extracted feature we remove outlier to remove the experimental error. To remove outlier we do clustering to check which record behavior does not belong to any cluster. Then we weight the attribute using information gain and information gain ratio to confirm which feature has capability to optimize result.

There are two kinds of feature we are extract from text, readability and style feature. Readability feature consist of Automated readability index, Coleman-liau, Flesh-Kincaid readability test, Gunning fox index, smog, flesh readability ease readability, lasbarhet index.

The readability feature measure of intelligibility that gauges the times of training expected to comprehend a bit of composing. These are generally utilized, especially to check wellbeing messages. It measures the level of people that are typing the text in English. These features are used in some published paper and show good results [22]. These results are the reason to use these feature in our paper. Standard formulas are globally available to estimate the readability feature.

Style feature consist of number of word, number of syllable, number of polysyllable, number of complex word, number of punctuation, number of word match, number of sentences. The attribute are also used in paper [22]. These are actually combination of length and relevance. These attribute also show good results in paper [22]. Word match and sentence match are showing the answers relevance with question. We avoid sentences match because it does not show good information gain. numbers of syllable, number of polysyllable, number of complex word, number of punctuation are showing the level of user, quality of content and manage length with above attribute.

Table 3. Selection of variables used with classifiers

| Readability Features | | | | |
|----------------------|---|-----------|--------|------------|
| S.no | Feature | Info gain | Status | References |
| 1 | Automated readability index | 0.130 | ✓ | [22] |
| 2 | Coleman-liau | 0.11 | ✓ | [22] |
| 3 | Flesch–Kincaid readability tests | 0.16 | ✓ | [22] |
| 4 | Gunning fox index | 0.10 | ✓ | [22] |
| 5 | Smog | 0.44 | ✓ | [22] |
| 6 | The Flesch Reading Ease Readability Formula | 0.25 | ✓ | [22] |
| 7 | Lasbarhets index | 0.12 | ✓ | [22] |
| Style Features | | | | |
| 8 | No of word | 0.38 | ✓ | New |
| 9 | No of paragraph | 0.07 | X | New |
| 10 | No of complex word | 0.21 | ✓ | New |
| 11 | No of syllabus per word | 0.23 | ✓ | New |
| 12 | No of poly syllabus | 0.24 | ✓ | New |
| 13 | No of punctuation | 0.05 | X | New |

| | | | | |
|----|----------------|------|---|------|
| 14 | Sentence match | 0.09 | X | [22] |
| 15 | Word match | 0.43 | ✓ | [22] |

In this table we see that there are 15 attributes, nine are baseline feature and 6 are proposed feature. We used 12 features and avoid four features because it contains less than 0.1 information gain. Readability features are basically shows the level of user like beginner, intermediate or professional. It also shows the quality of content. In readability smog give maximum information gain. In style feature we further divide three categories, length feature (no of word and no of sentences). Second use relevance feature, even every answer against a question is relevant, but we used to see that can user answer has same terms as asker used. If it used same terms it helps other to read easily. Third category used to check the quality of content.

Motivation of selecting readability features (Automated readability index, Coleman-liau, Flesh-Kincaid readability test, Gunning fox index, smog, flesh readability ease readability, lasbarhet index) is a readability test for English content, design to measure the understandability of content. It creates approximate representation of the US review level expected to comprehend the content. These features express the level of users by weigh the content. These features also have enough information gain mention in table to use in prediction. These features are used in some published paper and show good results [22], where it shows that if we rank content on text base then it performs better, and we are using text base features. We also used style features that are basically consist of relevance, length and quality. Motivation against using the Relevance feature does not means we are actually focus on relevance, if answer is given under question then it is understood it is relevant to answer[19]. It shows a user belong to specific field because different field has some specific terms, and if a user belong to that field than it use related terms. Other reasons, it has better information gain than other attributes used. Second part of style feature is consisting of length feature. Length feature has enough information gain to use, often answer consists of yes or no answers, maybe these answers has correct answer but these information or content are not enough to satisfy the asker, that is why answer are not select as best answer or rank. In paper “Belief and biases in web search” rank the answer “yes or no”, where it discuss that length feature effect the answer because it contains the more content and often select the best answer because it has maximum information and easily satisfy the asker. But quality of content compromise, we use quality feature to elaborate the length feature, it contains attribute: complex word, syllabus, poly-syllabus. These features have good information gain. It measures the quality of content. It helps the lengthy content to maintain its quality.

From above work we try to make our dataset is as gainful to our methodology that we can reach our maximum optimize result with help of these dataset.

B. Performances Evaluation

Normalized discounted cumulative gain (NDCG) method is a measure of ranking quality. It measures the gain of answers at their present position [25]. Actually we used NDCG to calculate the gain of answers at present position; it helps us to know which position of answer give us maximum gain. Then we can rank the other answer with the support of that position answers. Actually we obtain gain at different values of k (means different level of answers), it shows on which level of k we get maximum gain and useful to predict the data. Answers at k=1 is highest rank and decreasing gradually (k=2,3,..,10), in reality gain at k=1 will be maximum because it is high ranked, if not so there should be some problem in experiments. To observe this behavior we used the normalized discounted cumulative gain at top k (NDCG at top k). It is define as:

$$NDCG = \frac{1}{N} \sum_{i=1}^k \frac{2^{r_i}}{\log_2 (i+1)}$$

Where r_i is a true rating at position i in the ranking, N is the normalization factor [22].

C. Result and Discussion

Now we describe the estimation of baseline and proposed methodology. In this section, we further divide it in three parts. First we do comparison between Random Forest, Multiple Additive Regression Tree and Support Vector Machine with baseline features, we also compare Support Vector Machine with baseline and proposed features and finally comparison Random Forest, Multiple Additive Regression Tree and Support Vector Machine with combine features (baseline and proposed features).

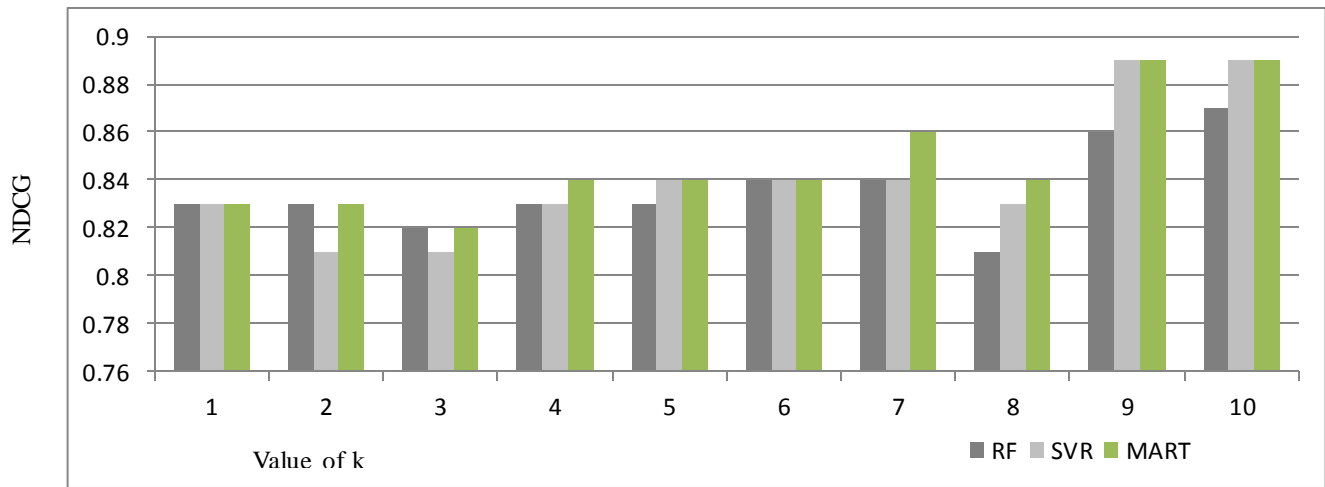


Figure 1. Comparison of RF, MART and SVR with baseline features

Table 4. Values of RF, MART and SVR with baseline features

| K | RF | SVR | MART |
|----|------|------|------|
| 1 | 0.83 | 0.83 | 0.83 |
| 2 | 0.83 | 0.82 | 0.83 |
| 3 | 0.82 | 0.81 | 0.82 |
| 4 | 0.83 | 0.83 | 0.84 |
| 5 | 0.83 | 0.84 | 0.84 |
| 6 | 0.84 | 0.84 | 0.84 |
| 7 | 0.84 | 0.84 | 0.86 |
| 8 | 0.81 | 0.83 | 0.84 |
| 9 | 0.86 | 0.89 | 0.89 |
| 10 | 0.87 | 0.89 | 0.89 |

In Figure 1, we calculate value according to k value. K=1 is the highest rank and decreasing accordingly. NDCG is actually calculating gain at different value of k. actually the gain at k=1 is high and decrease afterward, but we can see that the value is increasing. It will miss guide. In baseline paper, describe this problem as: all questions do not have at least 10 answers against a question. The results, I calculate with Yahoo Answers dataset is approximately matched with baseline paper and behave in same manner. Baseline feature are basically describe the only the level of Users, it is basically American standard to describe Users level in English writing. We can manage these value with add different features.

Table 4 shows gain obtained from highest rank is approximately same with all the classifiers and every question has one answer to rank minimum. In our analysis every question has minimum three answers against a question, and we can see that the gain against first three values is performing well. First answer give maximum gain and then decrease at second and third but with lower values of k, it does not perform well. Values at k=9,10 giving maximum gain, because there are less question that have nine and ten answers, minimum randomness give maximum gain. It is actually not behaving better.

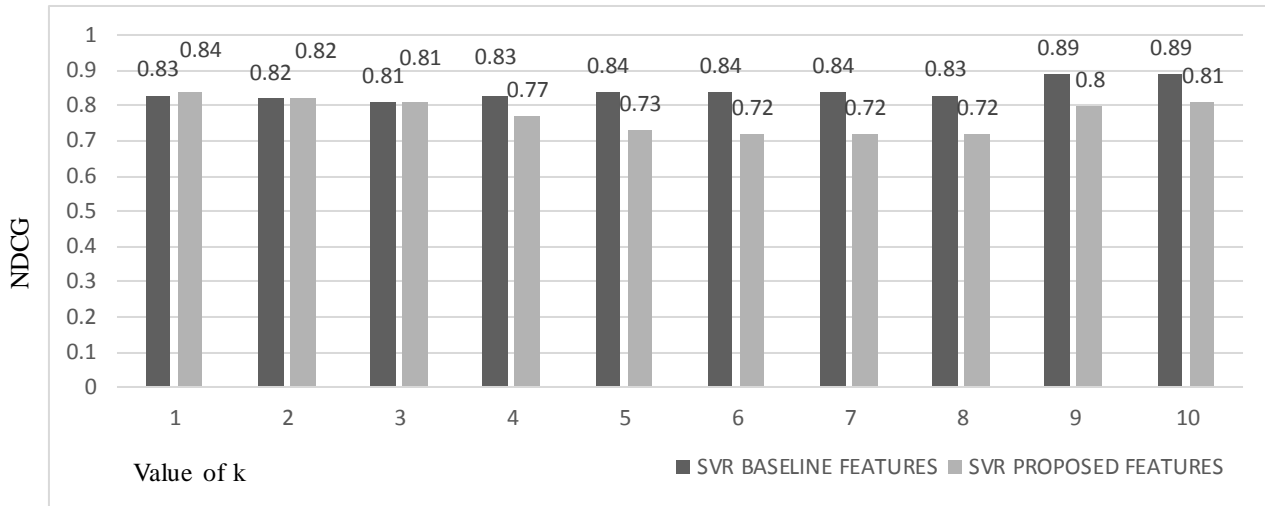


Figure 2. Comparison of SVR with baseline and proposed features

In Figure 2, we observe the proposed and baseline feature with SVR classifier. Proposed feature optimize result at $k=1$ of 0.01. Actually here length feature help us to optimize result because according to our observation often top rated answers have maximum description and proofs but we also maintain the content quality with quality feature. We can see that these features are also support us to maintain gain according to rank. $K=1$ is highest rank and decrease gradually with $k=2, \dots, 10$. We see in Figure 1, we analyze that every question has miming three answer and gain behave according to rank, but after three due to random number of answer gain miss-guide us. Proposed features help us to guide the actual gain at $k=4, \dots, 8$. But still at $k=9$ and 10, very less amount of answer given against question, but will improve this situation with combination of proposed and baseline features. Actually best gain given answers help us to predict the accurate result. We want to use those answers that has given good gain at different point of k . overall observation about proposed feature is: first to optimize results and second manage the gain when answers given in random number at different point of k .

Table 4. Values of RF, MART and SVR with combine features

| K | RF | SVR | MART |
|----|------|------|------|
| 1 | 0.82 | 0.84 | 0.83 |
| 2 | 0.80 | 0.83 | 0.82 |
| 3 | 0.78 | 0.81 | 0.80 |
| 4 | 0.76 | 0.79 | 0.78 |
| 5 | 0.76 | 0.77 | 0.76 |
| 6 | 0.76 | 0.77 | 0.76 |
| 7 | 0.74 | 0.76 | 0.73 |
| 8 | 0.70 | 0.75 | 0.71 |
| 9 | 0.78 | 0.77 | 0.77 |
| 10 | 0.80 | 0.77 | 0.79 |

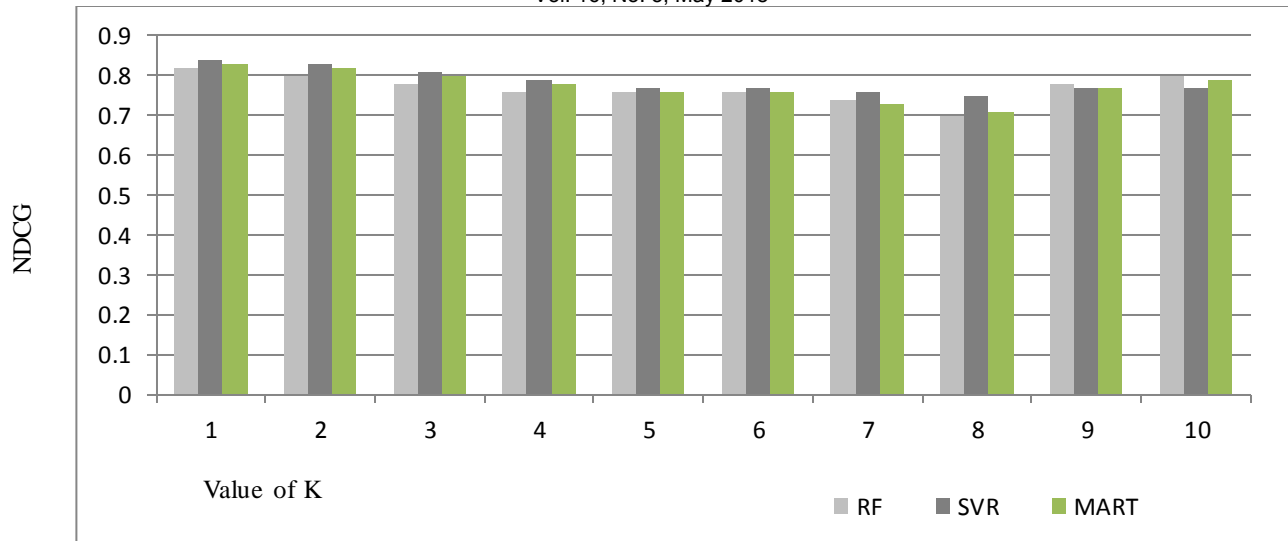


Figure 3. Comparison of RF, MART and SVR with combine features

In Figure 3, we can see that with the help of combine features (baseline and proposed features) overcome the problem of random number of against question, that create problem to calculate the accurate gain at point of k. in Figure 1 we observe that from k=4 to 10 gain obtain from low rank answer are high, but actually gain should be high at high ranked answers. In figure 2 we see that propose feature help us to overcome this situation at k=4 to 8, but still problem at k=9 and 10. There are two reason behind situation is these are low rank and we try to avoid low rank to predict, often only high rated answers are used. But with SVR classifier and combine feature we can get approximately better result, here length and quality features are basically helps us to manage these result and in figure 10 further results behave according to rank. We come to know that high ranked answers are given better gain than low rank.

VI. CONCLUSION

In this work, we proposed SVR approach to rank answers in Yahoo Answers. We adopted Approach of SVR and use two groups of features. These features are text features named readability and style feature. There are 12 attribute in both features. We used NDCG to calculate gain of from SVR.

Previous methodology used text feature with RF, result are worthy with stack flow dataset. But on yahoo dataset it does not show good result on yahoo dataset than before. So, we used SVR on yahoo dataset that give better result than RF give wit yahoo dataset.

Finally we compare result features individually, to check the behavior of feature, we come to know that style feature increase accuracy in result than readability feature.

In future, we further extend this work and use textual and non-textual feature to check the result in combination and individual. I hope these combinations will give better result.

VII. REFERENCES

- [1] White, R. (2013). Beliefs and biases in web search. In Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval (pp. 3-12).
- [2] Montazi, S., Klakow, D. (2010). Yahoo! answers for sentence retrieval in question answering. In The Workshop Programme (p. 28).

- [3] Jiang, J., Hassan Awadallah, A., Shi, X., & White, R. W. (2015). Understanding and predicting graded search satisfaction. In Proceedings of the Eighth ACM International Conference on Web Search and Data Mining (pp. 57-66)..
- [4] Agichtein, E., Brill, E., and Dumais, S. (2006). Improving web search ranking by incorporating user behavior information. Proc. SIGIR, 19–26.
- [5] Witten, I. H., & Frank, E. (2005). Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann.
- [7] BiologicalM. (2016). Biases .Retrieved From <https://en.wikipedia.org/wiki/bias> (date accessed: 2016, October 2).
- [8] Keiiri. (2015). Yahoo Answers. Retrieved From <https://en.wikipedia.org/wiki/yahooanswers> (date accessed: 2016, September 12).
- [9] Kimcapps. (2016). Yahoo dataset. Retrieved from <https://webscope.sandbox.yahoo.com> (date accessed : 2016, October 12)
- [11] Frakes, W. B., & Baeza-Yates, R. (1992). Information retrieval: data structures and algorithms.
- [12] Kayhan, V. (2015). Confirmation Bias: Roles of Search Engines and Search Contexts.
- [14] Raju, c. (2016). Retrieved From <https://en.wikipedia.org/wiki/informationreterival> (date accessed: 2016, October 2).
- [15] White, R. W., & Hassan, A. (2014). Content bias in online health search.ACM Transactions on the Web (TWEB), 8(4), 25.
- [16] White, R. W. (2014). Belief dynamics in Web search. Journal of the Association for Information Science and Technology, 65(11), 2165-2178.
- [17] Liu, X., Fang, H., & Cai, D. (2015). Towards Less Biased Web Search. In Proceedings of the 2015 International Conference on The Theory of Information Retrieval (pp. 373-376). ACM.
- [18] Livshits, B., & Mytkowicz, T. (2015). Interpol: Crowd-Sourced Internet Polls. In LIPIcs-Leibniz International Proceedings in Informatics (Vol. 32). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
- [19] Yan, d. Tang, j. (2016). Ranking and relevance in yahoo. Retrieved From

- [20] Kimcapps. (2016) .Yahoo answers. Retrieved from <https://answers.yahoo.com/> [\(date accessed: 2016, September 19\).](#)
- [22] Denil, Hassan. (2013). Exploiting User Feedback to learn to rank answer in Q&A Fourm. Proc. SIGIR'13, July 28–August 1, 2013.
- [23] kilash, c. (2016). Retrieved From <https://en.wikipedia.org/wiki/Svr> [\(date accessed: 2017, October 2\).](#)
- [24] david, c. (2017). Retrieved From <https://en.wikipedia.org/wiki/Randomforest> [\(date accessed: 2017, December 19\).](#)
- [25] Khanihal, j. (2017). Retrieved From [https://en.wikipedia.org/wiki/ Discounted_cumulative_gain](https://en.wikipedia.org/wiki/Discounted_cumulative_gain) [\(date accessed: 2018, January 12\).](#)
- [26] Argreich, L. (2014). Retrieved From <https://www.quora.com/When-can-we-use-support-vector-machine-basic-over-other-classification-algorithms> [\(date accessed: 2018, January 29\).](#)
- [27] Raczek, E., & Zagajewski, B. (2017). Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. European Journal of Remote Sensing, 50(1), 144-154.
- [28] Hasan, M. A. M., Nasser, M., Pal, B., & Ahmad, S. (2014). Support vector machine and random forest modeling for intrusion detection system (IDS). Journal of Intelligent Learning Systems and Applications, 6(01), 45.
- [29] User, T. (2017). Retrieved From https://en.wikipedia.org/wiki/Precision_and_recall [\(date accessed: 2018, March 1\).](#)
- [30] Kritikopoulos, A., Sideri, M., & Varlamis, I. (2007, May). Success Index: Measuring the efficiency of search engines using implicit user feedback. In the 11th Pan-Hellenic Conference on Informatics, Special Session on Web Search and Mining.

ACSR Overhead Transmission lines Modeling Considering Sag-Tension Estimation in Cold Area with different Wind Speeds

Muhammad Zulqarnain Abbasi & Hamza Umar Afridi

IQRA National University, Pakistan

Abstract—Load centers get generated electricity from power stations that are usually far; uninterrupted consumption or usage of power has increased in last few years. Transmission system is the system by means of which electricity is transferred from place of generation to the consumers. Overhead wires or conductors are the medium used for transmission of power. These wires are visible to wind, heat and ice. The efficiency of the power system increases if the losses of these overhead wires are minimal. These losses are based on the resistive, magnetic and capacitive nature of the conductor. It is necessary to create or make proper design of these conductors accompanied by proper installation. To balance the working and strength of overhead transmission line and to minimize its capacitive effect the conductors must be installed in catenary shape. The sag is required in transmission line for conductor suspension. The conductors are appended between two overhead towers with ideal estimation of sag. It is because of keeping conductor safety from inordinate tension. To permit safe tension in the conductor, conductors are not completely extended; rather they are allowed to have sag. For same level supports this paper provides sag and tension estimation with different wind speeds under low operating temperature 2°C . To calculate sag-tension estimation of ACSR (Aluminum Conductor Steel Reinforced) overhead lines three different cases are provided with normal and high wind speed effects. Four different span lengths are taken for equal level supports. ETAP (Electrical Transient and Analysis Program) is used for simulation setup. The results shows that wind speed has great impact upon line tension and with addition of wind speed the sag of line remains unaltered while tension changes. Moreover tension gets increase while increase in wind speed.

Keywords—component;ACSR;Span;Sag;Tension

I. INTRODUCTION

The power system comprises of three sections Generation, Transmission and Distribution. In this manner, from generation to the distribution, the electrical power is transmit through conductor materials. These conductor materials are overhead or underground. Both have its own particular pros and cons yet the majority of the conductor materials are overhead. The transmission lines are the significant part of any transmission network. These transmission lines convey electrical power over

long separations from nation one end to another end and in some cases nations to nations [1]. Because of this significance of transmission lines there suitable modeling is exceptionally essential. The execution of these transmission lines relies upon their appropriate modeling. Thusly, appropriate modeling of these transmission lines are one of the significant issue while raising and planning of transmission system [2].

A transmission system comprises of conductors, insulators and towers. Among these, conductors play a vital role because power flow through the conductor. Various types of conductors are utilized for the transmission of electrical vitality e.g. ACSR (Aluminum Conductor Steel Reinforced), AAAC (All Aluminum Alloy Conductor), AAC (All Aluminum Conductor) and HTLS (High Temperature low Sag). Among these all conductors, the ACSR conductor has certain advantages. ACSR have the galvanized steel core that carries the mechanical load and the high immaculateness aluminum, which carries the current. They use the lower thermal development coefficient of steel contrasted with aluminum-based conductors AAC and AAAC. These conductors have likewise great strength because of steel [3].

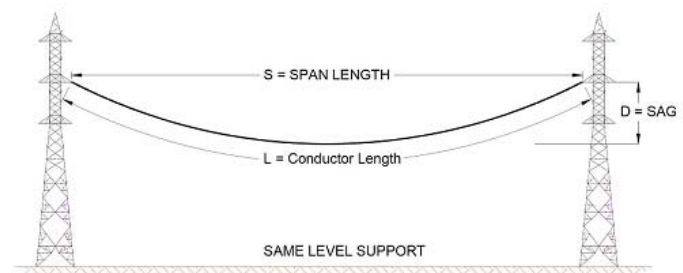


Figure 1. Sag at Same Level Support

The above figure illustrates that two overhead transmission line towers which are placed at same level. While point 'D' is distance between point of support and the lowest point on conductor that is referred as sag and point 'S' showing the distance between two towers that is named as span length.

II. OVERHEAD POSITION ANALYSIS

As appeared in the beneath Figure 2, a transmission line is connect at point A and B of two equal towers in a hanging shape. The point A and B are at equivalent distance from the ground. In this way as indicated by our significance of Sag, level difference of point A and B and most minimal point O is referred as sag [4.]

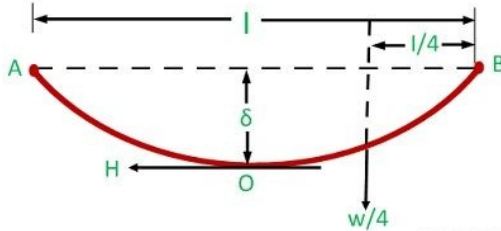


Figure 2. Overhead Position Analysis

The sag is critical within the transmission line. While designing an overhead transmission line, it is in need to focus that conductors have safe tension. If the conductors are unnecessarily extend out between two points of towers to save conductor material, at that point it may happen that tension of conductor accomplishes an appraised risky esteem and conductor break will happen [4].

In this way, to have safe tension in conductor, the conductors ought not to be extend excessively rather a sufficient sag or dip in transmission line is given [4]. The sag or dip in transmission line is given to keep the tension in the conductor inside the shielded and motivator in case of variety in tension in the conductor on account of seasonal variation. The sag varies with tower position. The tower set on plain surface are at same level, so it is easy to accomplish safe sag, tension and ground clearance level but due to wind effects sometime it is difficult to maintain sag-tension and ground clearance within safe limits. While in sloping territories, these supports are not usually located at same level, so the sag, tension and ground level varies. The ground level varies in hilly regions. So the sag also does not stay steady [6].

III. ACSR (ALUMINUM CONDUCTOR STEEL REINFORCED)

ACSR, a standard of the electrical utility industry since the mid 1900's, involves a solid or stranded steel center included by no less than one layers of strands of 1350 aluminum. Truly, the steel sum used to get higher quality soon extended to a significant section of the cross-portion of the ACSR, yet more starting late, as conductors have ended up being greater, the pattern has been to less steel content. To meet evolving necessities, ACSR is available in a broad assortment of steel substance - from 7% by weight for the 36/1 stranding to 40% for the 30/7 stranding. Early outlines of ACSR, for instance, 6/1, 30/7, 30/19, 54/19 and 54/7 stranding included high steel content, 26% to 40%, with accentuation on quality possibly in light of fears of vibration exhaustion issues. Today, greater sizes, the most used standing's are 18/1, 45/7, 72/7, and 84/19, include an extent of steel substance from 11% to 18%. For the unobtrusively higher quality 54/19, 54/7, and 26/7 standing's,

the steel substance is 26%, 26% and 31%, separately. The higher quality ACSR 8/1, 12/7 and 16/19 standing's, are used generally for overhead wires, extra long traverses, stream crossing points and so on [5].

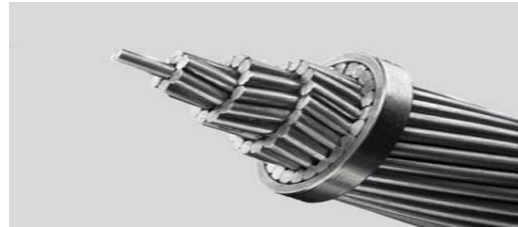


Figure 3. ACSR (Aluminum Conductor Steel Reinforced)

IV. EFFECTS OF WIND ON SAG AND TENSION

A weight is put by wind upon conductors will raise the conductor observable weight that results increment in tension. The increase in tension will expand the length of line due to flexible expansion. This expansion in resultant load will achieve a sag in incline direction with vertical and horizontal segments. The maximum working tension usually occurs at the maximum wind and everyday ambient temperature. So line tension has influenced by wind. For this reason, we apply distinctive wind ranges I-e normal and maximum to analyze the wind affect on line sag and tension [7].

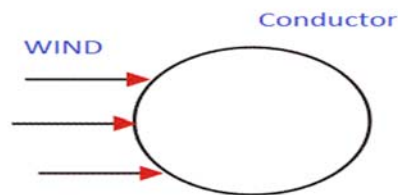


Figure 4. Direction of Wind Force on Conductor

In the above figure the wind is exerting a force on conductor due to which the apparent weight of conductor increase that results increase in tension of line. A wind on the conductor will expand the evident weight of the conductor bringing about an in increment in tension. This expansion will bring about a viable sag in a slanted heading with both even and vertical parts. To check its impact on sag and tension of ACSR overhead line simulation and analysis are performed for different wind speeds [6].

V. METHODOLOGY

ACSR conductor is chosen for simulation to check the sag and tension under various operating conditions in light of the fact that ACSR have the galvanized steel core that carries the mechanical load and the high immaculateness aluminum carries the current and use the lower thermal expansion coefficient of steel.

ETAP 12.6 programming is utilized for the estimation of sag and pressure of transmission line. The ETAP is among the best programming for electric power framework designing, planning and operation. ETAP Transmission and Distribution Line Sag and Tension module is an imperative tool to perform sag and tension estimation for transmission and distribution lines to ensure adequate operating condition for the lines. [9]. It is the minimal effort accessible programming for the count of sag and tension of various conductors. For simulation setup, we have considered equal level spans with towers height 20m. The configuration of conductors is set as horizontal and the spacing between the conductors is 1.5m. We have considered three different cases i.e. Case A, B and C. In case A sag-tension of ACSR is analyzed under low operating temperature i.e. 2°C with no wind effect because in winters usually the temperature is low. While in Case B the temperature is same as it is in the previous case but with the addition of normal wind speed i.e. 30 N/m². In Case C we considered low operating temperature with maximum wind speed 60 N/m².

The main reasons to choose ACSR for this research work are:

- ACSR is a type of high-limit, high-quality stranded conductor regularly utilized as a part of overhead electrical lines.
- The external strands are high-immaculateness aluminum; chosen for its great conductivity, low weight and cheap cost.
- Moreover conductor sag is less and a breakdown chance of conductor reduces.
- The center strand is steel for extra strength to help support the weight of the conductor.

VI. RESULTS AND DISCUSSION

A. Case 1

In Case 1, sag-tension is analyzed under low operating temperature i.e. 2°C because in winters the temperature fall down due to decrease in temperature the overhead lines contracts as a result there will be low sag. In table A for same level supports four different span lengths in minimum operating temperature i.e. 2°C are analyzed using ACSR.

TABLE I. LOW OPERATING TEMPERATURE WITH NO WIND

| Type of Conductor | Span(m) | Wind Speed N/m ² |
|-------------------|---------|-----------------------------|
| ACSR | 100 | 0 |
| | 200 | 0 |
| | 300 | 0 |
| | 400 | 0 |

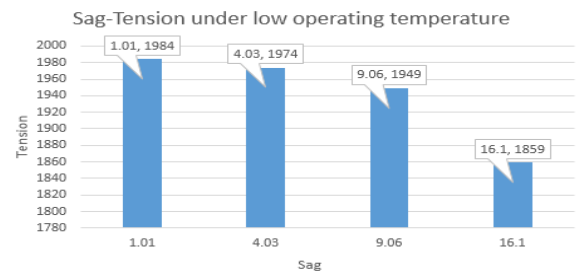


Figure 5. Sag-Tension Results with No Wind

When the length of span is 100m the sag is 1.01m and tension is 1984. As the length of span increases from 100m to 200m the sag is 4.03m and tension is 1974. Similarly for 300m and 400m the sag is 9.06, 16.1 while tension is 1949 and 1859 respectively. From the above figure it is shown that as the length of span increases the sag likewise increases this is because sag is directly proportional to length of span and inversely proportional to tension.

B. Case 2

In this case, low operating temperature with normal wind speed i.e. 30 N/m² is analyzed. As the table below showing low operating temperature with normal wind speed for different span lengths are analyzed.

TABLE II. LOW OPERATING TEMPERATURE WITH NORMAL WIND SPEED

| Type of Conductor | Span(m) | Wind Speed N/m ² |
|-------------------|---------|-----------------------------|
| ACSR | 100 | 30 |
| | 200 | 30 |
| | 300 | 30 |
| | 400 | 30 |

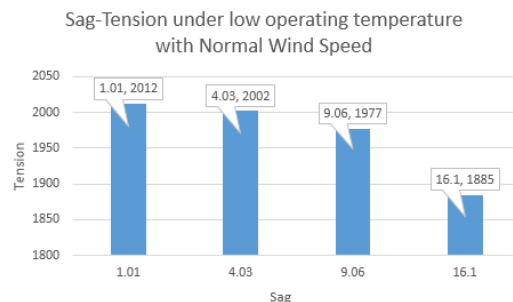


Figure 6. Sag-Tension Results with Normal Wind Speed

From the above graph when the length of span is 100m the sag is 1.01 and tension is 2012. As the span length increased i.e. 200m the sag is 4.03 and tension is 2002. Moreover for 300m & 400m the sag is 9.06, 16.1 and tension is 1977 and 1885 respectively. The figure shows that with the addition of wind the tension of the line increases while the sag remains unaltered. This is because wind applies a force upon the conductor as a result apparent weight of conductor increases that increase tension.

C. Case 3

In Case 3, sag-tension is analyzed under low operating temperature with high wind speed i.e. 60 N/m². Because sometimes the speed of wind is higher than normal days and also the wind load on conductor increase the apparent weight of conductor due to which there is increase in tension. Therefore in this case we have considered maximum wind speed. In Table B below, Low operating temperature with high wind speed is analyzed for different spans length using ACSR.

TABLE III. LOW OPERATING TEMPERATURE WITH HIGH WIND SPEED

| Type of Conductor | Span(m) | Wind Speed N/m ² |
|-------------------|---------|-----------------------------|
| ACSR | 100 | 60 |
| | 200 | 60 |
| | 300 | 60 |
| | 400 | 60 |

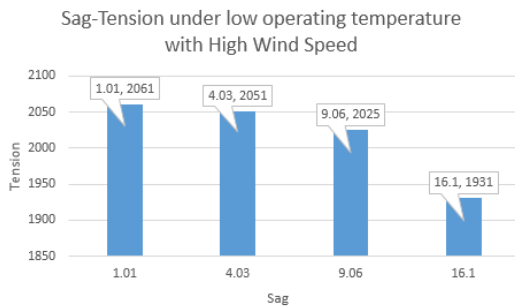


Figure 7. Sag-Tension Results with High Wind Speed

From the figure it shows that when the length of span is 100m the sag is 1.01 and tension is 2061. While for 200m span length sag is 4.03 and tension is 2051. Similarly for 300m & 400m span sag is 9.06, 16.1 and tension is 2025 and 1931 respectively. With the increase in wind speed in this case, the sag remains unaltered as in previous case while tension with the high wind speed further increases. The graph shows that tension gets increase with increase in wind speed.

VII. CONCLUSION

In this research paper, different types of wind speed are discussed under low operating temperature. Three different cases are analyzed for sag-tension estimation using ACSR overhead lines. Four different span lengths are selected which are at same level support. After simulation results following conclusions are made:

- In cold weather, the temperature is usually low i.e. less than 2 °C due to decrease in temperature the overhead lines contracts that results in low sag

which will indicate high tension. The length of span increases the sag likewise increases this is because sag is directly proportional to length of span and inversely proportional to tension.

- With the addition of wind in cold weather, the sag of overhead lines remains unaltered as in previous case but with the addition of wind the tension of line increases this is due to wind applies a force upon conductor as a result apparent weight of conductor increases that increases tension.
- With maximum wind speed in cold weather, the sag remains unchanged as it was in previous case but tension with the high speed of wind further increases because wind speed has great impact upon line tension as tension get increase while increase in wind speed.

From this paper, one can undoubtedly discover the sag-tension estimation of ACSR overhead lines for distinctive instances of wind speed without calculating it mathematically.

Accordingly for overhead transmission lines sag tension estimation, ETAP tool is exceptionally useful to anticipate sag-tension behavior for overhead transmission lines. Moreover it is effectively accessible and cheap software for calculating sag-tension estimation as compared to expensive commercial software.

REFERENCES

- [1] Chaudhari Tushar, Jaynarayan Maheshwari and Co. 'Design and Reconductoring of A 400 K.V Transmission Line And Analysis on ETAP'. International Journal of Engineering Research and Development (IJERD) ISSN: 2278-067X Recent trends in Electrical and Electronics & Communication Engineering (RTEECE 17th – 18th April 2015).
- [2] I. Zamora, A.J. Mazon and Co. High-temperature conductors: a solution in the uprating of overhead transmission lines. Power Tech Proceedings, 2009 IEEE Portugal.
- [3] Sag-Tension Calculation Methods for Overhead Lines (2007). CIGRE B2-12 Brochure (Ref. No. 324) pp. 31-43.
- [4] V.K. Mehta and Rohit Mehta (2014). Principles Power System. S. Chand and Company Pvt. Ltd. Ram Nagar New Delhi.
- [5] I. Albizu, A. J. Mazon, and E. Fernandez (2011). "A method for the Sag-tension, calculation in electrical overhead lines. International Review of Electrical Engineering, volume 6, No. 3 pp. 1380-1389
- [6] Oluwajobi F. I., Ale O. S. and Ariyanninuola A (2012). Effect of Sag on Transmission Line. Journal of Emerging Trends in Engineering and Applied Sciences (JETEAS) 3 (4): 627-630 © Scholar link Research Institute Journals, (ISSN: 2141-7016).
- [7] M. Zulqarnain Abbasi (2018) Sag-Tension Analysis of AAAC Overhead Transmission lines for Hilly Areas (IJCSIS) April 2018, Vol. 16 No. 4
- [8] S. Kamboj R. Dahiya (2014) Case Study to Estimate Sag in Overhead Conductors Using GPS to Observe the Effect of Span Length. 978-1-4799-3656-4/14/\$31.00 2014 IEEE.

Comparison of Machine Learning Techniques for Identification of Disease

Shaveta Malik¹, Tapas Kumar², A.K Sahoo^{3*}

1, 2 School of Computer Science, Lingaya's University, Faridabad, Haryana, Shavetamalik687@gmail.com

3* Scientist, Biodiversity Lab, Reverie Ecology and Fisheries Division, Barrackpore, Kolkata

Abstract:-The Fish disease causes several losses in the fish farm. A number of fungal and bacterial diseases especially EUS (Epizootic Ulcerative Syndrome) causing Morbidity and Mortality in fish. Fish Infection caused by *Aphanomyces invadans* is commonly known as EUS (Epizootic Ulcerative Syndrome) and it is due to the primary fungal pathogen. EUS disease is still misidentified by the people. The paper proposed a combination of feature extractor with PCA (Principle component analysis) and classifier for better accuracy. Proposed combination of techniques gives better accuracy while identification of EUS and Non EUS infected fishes. After experimentation, it is found that PCA improves the performance of classifier after reducing the dimensions. Real images of EUS infected fishes have been used throughout and all the work is done in MATLAB.

Keywords: - *Epizootic Ulcerative syndrome (EUS), Principle component analysis (PCA), Features from Accelerated Segment Test (FAST), Neural Network (NN)*

I. Introduction

Fish is a dependable source of animal protein in evolving countries like INDIA. Due to large scale of mortality occurs among the fresh water fishes, it causes a immense loss to the nation. Spreading of EUS is a semi-global problem among the fishes of fresh water, in large natural water bodies may not be possible to control of EUS, and Control of EUS in large natural water bodies may not be possible. Today's major problem is to control and treatment of EUS. The accuracy of the final diagnosis found using experiences of fish farmers or fish veterinarian. Traditionally, Skills and experiences and the time spend by the individual defines the accuracy of the final diagnosis. Normally infected fish will die quickly if correct and accurate treatment is not provided. In order to solve this problem, combination of Feature extractor with PCA (Principle component analysis) applied to extract the feature and classifier applied to classify the EUS infected and Non-EUS infected fish in order to find the accuracy rate of EUS and Non-EUS infected fish. The infected fish will normally die very quickly if correct and effective treatment is not provided in time. Mortality of fish will affect the loss of fish farmers, Indian Market loss and automatically it will also affect the international market loss. The paper compares the combination of different feature extractor with different classifier for finding the accuracy .It finds that the proposed combination gives better accuracy. The accuracy has been found with the combination of Feature Extractor and PCA (Principle component analysis) and feature Extractor without PCA. The dimensionality reduction can be possible through PCA of the dataset and removes the dimensions which have the least important information. . The data utilizes less space if number of dimensions has been reducing, it helps in classification of larger dataset s in less time. In the classification experimentation, two classifier or classification algorithms have been taken to find the accuracy i.e. KNN (K-Nearest Neighbour) and Neural Network. PCA has been applied after extracting the feature from HOG (Histogram of Gradients) and FAST (Features from Accelerated Segment Test) of each image. It has been observed through results that PCA (Principle component analysis) improves the accuracy of classification. Many Researchers have done lot of work in many techniques related to feature extraction and area related to the paper. Jeyanthi Suresh et al.[1] In the paper, proposed a method or technique which automatically recognized the activity of human from the video with the feature extractor which was the HOG & Probabilistic Neural network (PNN) classifier. The classifier was used for classifying the actions of video experiments and results were found on Kth database and gave better performance, 89.8% accuracy for test and 100% for the training set and measured the performance of each featured set with different classifier..Valentin Lyubchenko et al. [2] in the paper selected the markers of colors to distinguished the infected and Normal area, there was a drawback in the methodology of false point which can be appeared as a disease area due to automatic allocation of color, it has the ability to change the marker while selecting the color in the segmented image. Hitesh Chakravorty et al. [3] suggested a method in which disease fish image recognized by using dimension reduction technique that was through PCA method and segmentation of fish image with K-means clustering technique, segmentation was based on the color features HSV images and Morphological operations for the area that is diseased its detection and dimensions. In which only handpicked EUS diseased images of the fishes were considered, the proposed method or technique to improve the diseased identification with larger accuracy as well as correctly detected diseased area. In which extracted the features and PCA applied which is principle component analysis and converted into feature vector Euclidian distance has been applied for classification.

II. Methodology

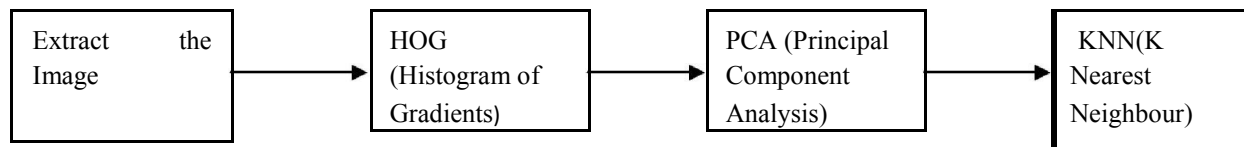


Figure 1a). Flow Chart of the Process through K-NN

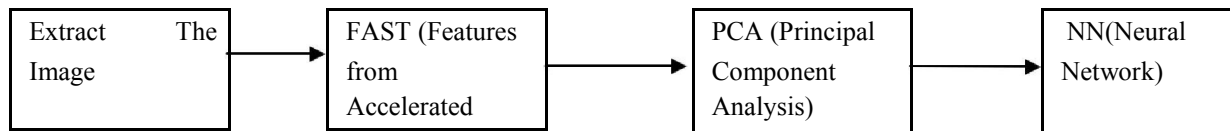


Figure 1b). Flow Chart of the Process Neural Network

In Figure 1a) shows the Flow chart of Process and steps applied to extract the features and find the performance of classification through K-NN classifier.

In Figure 1b) shows the Flow chart of Process and steps applied to extract the features and find the performance of classification through NN (Neural Network) classifier

The processes are broadly separated into the Four stages: - Pre-processing, Feature Extraction, dimensionality reduction and classification.

Stage 1:- Pre-processing- Real Images have been collected and remove the noise after that segmentation has been applied.

Stage 2:- Feature Extraction- In image processing extracting the features from the image, it is not possible to extract the feature from a single pixel, it interact with the neighbours also, feature extractor used to extract the feature from the image of EUS (Epizootic Ulcerative Syndrome) infected fish.

Stage 3:- Dimensionality Reduction:- After extracting the feature from HOG and FAST ,PCA (Principal Component Analysis) will apply for the dimension reduction of the features and amount of memory used by the data, It helps in faster classification also.

Stage 4:-Classification: - Classify the fish image into EUS Infected and Non-EUS infected through classifier e.g. KNN(K-Nearest Neighbor) and NN(Neural Network) and find the accuracy as dataset has EUS and Non-EUS infected fish image both.

2.1 HOG (Histogram of Gradients):-

It is based on the concept that divide the image into small area called cells and then form the blocks through cells e.g. 4*4 pixel size cell was selected by default and blocks size is 8*8 then Calculate the edge gradients e. g from each of the local cells 8 orientations are calculated and form the histogram of cell then normalize it and normalize the blocks also, small changes are done in the position of window in order to not to see the descriptor changing heavily and to get the lesser impact far from centre gradients of the descriptors. For each pixel in order to assign magnitude weight one half of the width of descriptor known as sigma is assigned

HOG Steps (Histogram of Gradients) in Matlab

Implementation Step1:- Input the image of EUS infected fish.



Figure 2: Input Image

Step 2:- Normalize the image or gamma which is the square root of image intensity depends on what kind of the image.

Step 3:- Orientation of gradient and its magnitude is computed.

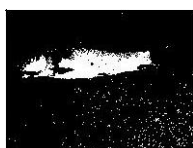


Figure 3: Gradient computed Image

$$\text{Gradient Magnitude } |\nabla f(x,y)| = \sqrt{f_x^2 + f_y^2} \quad (1)$$

$$\text{Gradient Direction } \theta = \tan^{-1} \frac{f_x}{f_y} \quad (2)$$

Where:

f_x is the derivative w.r.t x (gradient in the x direction)
 f_y is the derivative w.r.t y (gradient in the y direction).

Step 4:- Create and split the window into cells and each cell represents the pixels and make the histogram of orientation gradient.

Step 5:- Grouping the cell together into large and then normalize it.

Step6:- After extracting feature from HOG apply the Machine learning algorithm or classifier.



Figure 4: HOG descriptor Image

In Figure 4 Applied the HOG (Histogram of Gradients) to extract the features and then evaluated the performance of classification through Machine Learning algorithm.

2.2 FAST (Features from Accelerated Segment Test):-

Fast technique recognizes the interest point in an image basis intensity of local neighbourhood. It is the fastest and better algorithm than others, the identification of corners has been given priority over the edges[8], because they claimed that the corners have the most innumerable features which show a strong two-dimensional intensity change, and therefore the neighbouring points as well as the work of the algorithm, it makes pixels comparable to a fixed radii circle and to classify a point as a corner if a circle with maximum numbers of pixels on its radii can be drawn which are brighter or darker than its central point. The detector's main limitation here of is that almost all the features are closer to each other.

In figures 5 shows the original image and then after applied the FAST



Figure 5 a):- Original Image

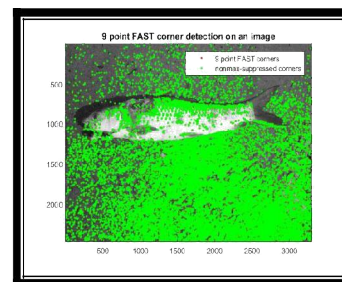


Figure 5 b):- FAST (Features from Accelerated Segment Test)

2.3 PCA (Principle component analysis):-After extracting features from the HOG (Histogram of Gradients) and FAST (Features from Accelerated Segment Test), PCA applied as features reduced by the PCA because it is to reduce the dimensionality of the dataset and by reducing the number of dimensions, it utilizes less space. It helps in classification on large dataset as it takes less time. After reducing the feature space some noise and redundancies in the features are eliminated while reduce the dimensionality.

2.4 HOG-PCA & FAST-PCA: - Feature vector dimensionality reduction is the work of PCA. Then on to the extracted features we apply PCA a better accuracy is found in the case of FAST -PCA application then in HOG-PCA. When FAST-PCA applied with the Machine learning algorithm it gives 3.8% classification accuracy higher as compared to the HOG-PCA. (Result shows in Figure 9 and 11).

2.5 KNN: - It is a supervised learning algorithm and is usually used in machine learning methods. The best way to classify the feature vectors is basis the closest training. Being an easy and efficient that depends on the known samples it is an important non parametric classification approach, depending on the known samples, according to the approximate neighbours of K-nearest which classify and specifies a class label for unknown samples. $(x, F(x))$ are being stored as examples of training. Being an input in memory an n-dimensional vector (a_1, a_2, \dots, a_n) is termed as x and corresponding output is $F(x)$ that is classified basis its neighbours as per their size for classification,, the value of K-nearest has been chosen[18] if K-nearest = 1, the only to the class of its neighbours the object is assigned, the it can reduce the effect of noise on the major value classification of K-nearest, but can separate the boundaries between the classes.

KNN is classified into Testing and Training Phase for classification:-

Training phase:

- 1) Select the images for training phase.
- 2) After that training images will read.
- 3) Pre-process and resize the each image.
- 4) Preprocessed image was used to extract the features (through HOG) to form a vector of features of image that are local to the image.
- 5) By the local features, feature vector is constructed of the image as row in a matrix.
- 6) Repeat steps 2 to Step 5 for all the training images.
- 7) Trained the KNN technique for the phase of testing.

- 1) Read the images for test.
- 2) After applied the KNN first, identified the nearest neighbours using the function of Euclidean distance by the Training data.
- 3) If the K neighbours have all the same labels, the image is labelled and exit otherwise compute pair wise distances between the K neighbours and construct the distance matrix.

III. Proposed Methodology:-

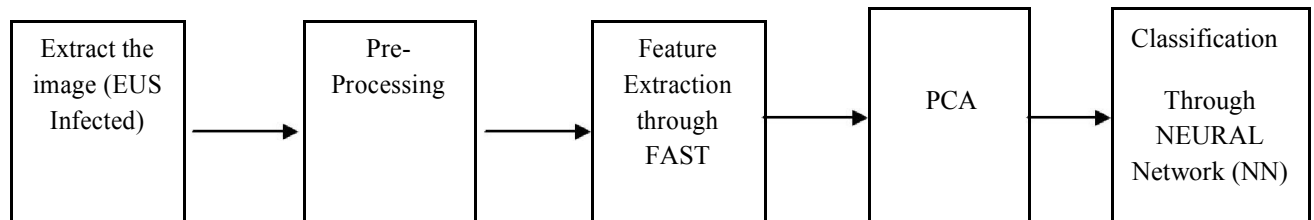


Figure 6:- Proposed Flow chart

In Figure 6, it shows the Proposed Flow chart or steps to be implemented to extract the features and find the classification performance through Machine Learning Algorithm (Neural Network)

EUS disease detection from the image, first apply the morphological operations i.e. The image is converted into greyscale and enhances the image, remove the noise and segmentation applied and then extract the feature from FAST then apply the PCA dimensionality reduction of the extracted features, match the features after applying the classifier which is neural network and find the classification accuracy.

The algorithm explained below of combination (FAST-PCA-NN) method:

1. A pixel is selected which is considered as “pe” in the image and assumed “IPE” the intensity of the pixel.
// Meaning of a pixel under test i.e., it is an interest or feature point which is to check.
2. T is taken as the threshold intensity set with the assumption that it will be around 19-20 percent of the available pixels.

3. IPE is assumed to be the pixel intensity of the 16 pixels of circle surrounding the pixel "pe".
((Bresenham circle [15] of radii 3.))
4. Threshold will distinguish the "N" pixels adjacent to the 16 pixel by checking if they are above or below it.
// (N = 12 as a part of the first form of the algorithm)
5. The intensity of pixels is comparing the 1, 5, 9 and 13 of the circle with IPE (Intensity of pixel)
first The algorithm that considered will be fast; it should be that no less than the three pixel combination should follow rule 4, so that the interest point will exist.
6. The pixel "pe" will not be considered as an interest point or corner in it that not less than the three of the above mentioned four pixel values I1, I5, I9, I13 are neither above nor below $I_{pe} + T$. hence, in such situations, pixel "pe" will be rejected from considering a corner point. Only if the least 14 least 3/4 the of the pixel are considered to be falling under that criteria.
7. Then the process will repeat for all image pixels.
8. After that Apply the PCA (Principal component Analysis) for reducing the dimensions.
9. After that applied the Neural Network to train and test the image
 - a) Take X as a variable and X= features (Input Data) // Extracted features from FAST Algorithm. [Input, Targets]=Datasets;
 - b) Create the Pattern Recognition Network
 - c) Then divide the data for Training ,Validation and Testing
 - d) Setup the data into training, validation.
 - e) Setup the division of data for Training ,Validation ,
 - f) Then the Network will be train and can be test the network after trained.

3.1 Sample of training Dataset:-



Figure 7:-Sample of Training Dataset (EUS Infected Fish)

In Figure 7 shows the Sample of EUS infected fish, The sample of EUS infected fish used in experimentation are the real images.

3.2 Performance Comparison between the combined Techniques

| Classification Accuracy | Percentage |
|-------------------------|------------|
| HOG-PCA-KNN | 56.32% |
| HOG-PCA-NN | 92.5% |
| FAST-PCA-KNN | 63.32% |
| FAST-PCA-NN | 96.3% |

Table: 1:- Comparison between classifications Accuracy of different combination Techniques

The Table 1 shows among all combination Technique, the Proposed Combination Technique shows 96.3% accuracy as it gives higher accuracy as compared to others in the paper.

3.2 a) Graph between All combinations:-

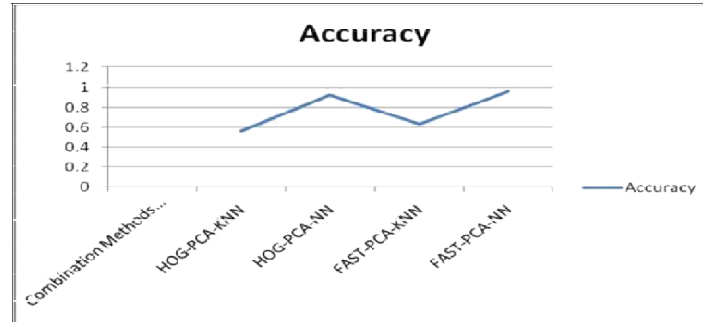


Figure 8:- Classification Accuracy between Combined Technique

In Figure 8 the Graph shows the Performance comparison between Existing and Proposed Combination of Technique.

$$\text{Performance Evaluation Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (3)$$

In Performance Evaluation Accuracy find the positive and negative rate to classify the EUS and Non-EUS infected fish.

3.3 Accuracy through HOG-PCA-NN and FAST-PCA-NN:-

After applying the Feature extraction through HOG and FAST and get the classification accuracy through Neural Network. It extract the 4356 features in order to get a neural network to successfully learn task, it must be trained first. The training database is then divided into testing set and training set. Neural network was trained using the train set. To get the better result train the neural network many times and get the average of classification accuracy. In which input or feature extracted by the feature extractor is 4356 and has taken 10 hidden layers which give the output. Testing set is used to test the neural network. To find the hidden neurons, in an architecture the dataset is partitioned into test and train data Ttrain sets Ttesting[16]. The test set is used to test the ability of the network [1]. Network pattern recognition is be implemented.

3.4 Results and Analysis:-

3.4 a) The Result shows the classification accuracy through HOG-PCA-NN

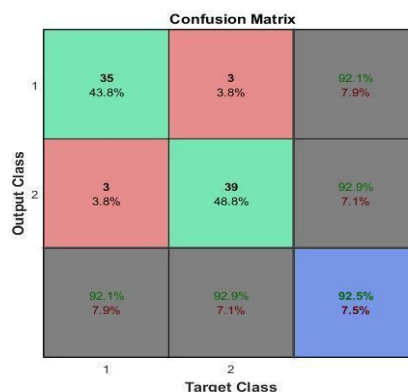


Figure:-9 Confusion Matrix (HOG-PCA-NN)

Figure 9 shows the confusion matrix of HOG-PCA-NN and gives the performance accuracy with Non-EUS and EUS fish.

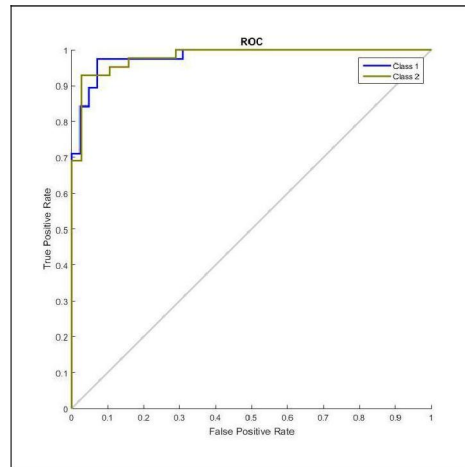


Figure:-10 Receiver Operating Characteristic Curve

In Figure 10 shows the ROC known as Receiver's Operating Characteristic curve of (HOG-PCA-NN) which gives the graph between True Positive Rate Vs False Positive Rate with the EUS and Non-EUS infected fish.

3.4 b):- FAST-PCA-NN

The results shows the classification accuracy through FAST-PCA-NN

| Confusion Matrix | | |
|------------------|-------------|--------------|
| Output Class | 1 | 2 |
| 1 | 37 46.3% | 2 2.5% |
| 2 | 1 1.3% | 40 50.0% |
| | | Target Class |
| | | 1 |
| | | 2 |
| | | 3 |

Figure 11:-Confusion Matrix (FAST-PCA-NN)

In Figure 11 shows the Confusion Matrix of (FAST-PCA-NN) which gives 96.3 % accuracy in correct detection of EUS disease fish and 3.7% not correctly classified ,the graph shows the Target class Vs output class, It tells the False positive(FP) and False Negative(FN), True positive(TP) and True Negative(TN).(Performance Accuracy)

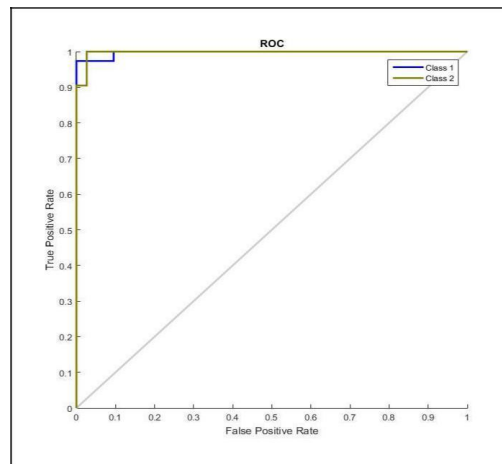


Figure 12:- Receiver Operating Characteristic Curve

In Figure 12 shows the ROC curve known as Receiver Operating Characteristic Curve (FAST-PCA-NN) shows the Receiver Operating Specification curve which gives the graph between True Positive Rate (TPR) Vs False Positive Rate (FPR) with the EUS and Non-EUS infected fish.

The ROC curve area shows the perfectly prediction when it comes 1 as said by the properties of ROC curve in Figure 12; an area of .5 represents a worthless detection or random prediction.

IV. Conclusion

The Experimental evaluation for performance Comparison shows the proposed combination (FAST-PCA-NN) gives better accuracy as compared to the other combinations in the paper. Proposed combination gives 3.8% better accuracy than other (HOG-PCA-NN) when it combines with PCA because it reduces the dimensionality of the dataset by reducing the number of dimensions. The Experimentation has been done on MATLAB Environment and on real images of EUS Infected.

Acknowledgement

Collection of the EUS infected fish's images have been done from National Bureau of Fish Genetic Resources (NBFGR, Lucknow) and ICAR-(CIFRI), Kolkata. Thanks to Dr. A.K Sahoo (CIFRI, Kolkata) and Dr .P.K Pradhan (NBFGR, Lucknow).

References

- [1] A. Jeyanthi Suresh , P.Asha , "Human Action Recognition in Video using Histogram of Oriented Gradient (HOG) Features and Probabilistic Neural Network (PNN)", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 7, July 2016
- [2]. Valentin Lyubchenko, Rami Matarneh, Oleg Kobylin, Vyacheslav Lyashenko, "Digital Image Processing Techniques for Detection and Diagnosis of Fish Diseases", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 7, July 2016
- [3]. Hitesh Chakravorty, Rituraj Paul & Prodipto Das , "Image Processing Technique To Detect Fish Disease", International Journal of Computer Science & Security (IJCSS), Volume (9) : Issue (2) : 2015 121
- [4]. Navneet Dalal, Bill Triggs, "Histograms of Oriented Gradients for Human Detection" International Conference on Computer Vision & Pattern Recognition (CVPR '05), Jun 2005, San Diego, United States.
- [5]. E. Rosten, R. Porter, T. Drummond, "FASTER and better: a machine learning approach to corner detection "IEEE Trans Pattern Anal Mach Intel, 32 (2010), pp. 105–119

- [6]. Jeong-Seon Park, Myung-Joo Oh, and Soonhee Han, “*Fish Disease Diagnosis System Based on Image Processing of Pathogens Microscopic Images*”, Frontiers in the Convergence of Bioscience and Information Technologies 2007
- [7]. Mikolajczyk, Schmid, “*D. Scale and affine invariant interest point detectors*”, International Journal of Computer Vision, 2008, vol. 60, no. 1, p. 63-86
- [8]. Muralidharan, R. Chandrasekar, C. “*Combining local and global feature for object recognition using SVM-KNN*”. In Proceedings of International Conference on Pattern Recognition. Informatics and Medical Engineering. 2012, ISBN 978-1-4673-1039-0/12.
- [9]. Rosten, E., & Drummond, T. (2006). “*Machine Learning for High-speed Corner Detection*”. Proceedings of the 9th European Conference on Computer Vision - Volume Part I (pp. 430-443). Berlin, Heidelberg: Springer-Verlag
- [10]. Burge, C. A., Mark Eakin, C., Friedman, C. S., Froelich, B., Hershberger, P. K., Hofmann, E. E., Ford, S. E. (2014). Burge, C. A., Mark Eakin, C., Friedman, C. S., Froelich, B., Hershberger, P. K., Hofmann, E. E., Ford, S. E. “*Climate change influences on marine infectious diseases: implications for management and society*”. Annual review of marine science, 2015, 6, 249-277.
- [11]. Lafferty, K. D., Harvell, C. D., Conrad, J. M., Friedman, C. S., Kent, M. L., Kuris, A. M., Saksida, S. M. (2015), “*Infectious diseases affect marine fisheries and aquaculture economics*”, Annual review of marine science, 7, 471-496.
- [12]. Afferty, K. D., Harvell, C. D., Conrad, J. M., Friedman, C. S., Kent, M. L., Kuris, A. M., Saksida, S. M. , “*Infectious diseases affect marine fisheries and aquaculture economics.*”, Annual review of marine science, 2015, 7, 471-496.
- [13]. Narasimha-Iyer, H., et al., “*Automatic Identification of Retinal Arteries and Veins From Dual-Wavelength Images Using Structural and Functional Features. Biomedical Engineering*”, IEEE Transactions on, 2007. 54(8): p. 1427-1435.
- [14]. Daoliang Lia, Zetian Fua, Yanqing Duanb, “*Fish-Expert: a web-based expert system for fish disease diagnosis*”, Expert Systems with Applications 23 (2002) 311–320.
- [15]. A. Alahi, R. Ortiz and P. Vandergheynst, “*FREAK: Fast Retina Keypoint*,” IEEE Conference on Computer Vision and Pattern Recognition, 2012.
- [16]. J.-Y. Chang and J.-L. Chen, “*Automated facial expression recognition system using neural networks*”, Journal of the Chinese Institute of Engineers, vol. 24, no. 3, (2001), pp. 345-356.
- [17]. M. Rizon, M. F. Hashim, P. Saad, S. Yaacob, “*Face Recognition using Eigenfaces and Neural Networks*”, American Journal of Applied Sciences, vol. 2, no. 6, (2006), pp. 1872-1875. .
- [18] Li Li, ZHANG YanXia & ZHAO YongHeng *k-Nearest Neighbors for automated classification of celestial objects*” Science in China Series G: Physics, Mechanics & Astronomy, jul, 2008, vol. 51, no. 7, pp 916-922

SiDe ENABLED RELIABLE REPLICA OPTIMIZATION

Prof. Anjali Chandavale¹, Kartik Hedau², Akshay Ghorpade³, Yash Palwe⁴,

Ayushi Dantulwar⁵

Abstract- Two most common concerns for current Cloud storage systems are data reliability and storage costs. To ensure data reliability, typically there is multi-replica (i.e. three replicas) replication strategy in current clouds is used which requires a large amount of storage consumption, that results in high storage cost for applications generating vast amount of data particularly in the Cloud. This paper presents a cost-efficient and data reliable management mechanism named SiDe (Similarity Based Deduplication) for the private cloud storage systems of enterprises. While storing the file the parameters file_name, file_size, file_storage_duration are taken as input from the user. To minimize memory requirement, the file is divided into fixed sized chunks and stores only the unique chunks along with the code for regeneration. A compressed copy is then created for the input files stored for long duration. A key generation algorithm is then used to generate a unique key for ensuring security. The simulation indicates that, compared with the stereotypical three-replica strategy, SiDe can reduce around 81-84% of the Cloud storage space consumption for file size varying between 10KB to 300KB and files having duplicate data chunks, hence considerably reduces the cloud storage cost.

Keywords- Minimizing data replication, SiDe, Data Deduplication, data reliability, cost-effective storage, cloud computing.

I. INTRODUCTION

Cloud storage has numerous benefits such as accessibility, scalability, cost efficiency etc., due to which many users are using cloud storage to store and backup their valuable data. In the present scenario the data generation rates are increasing due to which is a tedious task for cloud storage providers to serve efficient storage. Deduplication is one of the leading techniques used by cloud storage providers, which helps in saving 70%-75% of storage ensuring data reliability and effectively making it cost-efficient. [1]

In current cloud storage platform, the most commonly used approach is data replication which provides data reliability assurance, preventing the probability of data loss by creating multiple replicas of the data. For instance, cloud storage systems like Amazon S3 [2], Hadoop Distributed File System (HDFS) and Google file system (GFS) that adopt data replication schemes where three replicas, i.e., three copies of data including the original data, are stored. Yet the further growth of Cloud data can cause obstruction in the enhancement of cloud data storage, because the three replica replication technique consumes too much extra storage space, thus increasing the huge storage cost.[3]

The new kind of storage that is gaining much attention in the current scenario is cloud storage. One side, the digital data is increasing, another side backup problem and disaster recovery are becoming critical for the data centers[4]. The three-quarter of digital information is redundant by a report of Microsoft research. This massive growth in storage environment is balanced by the concept of deduplication.[12] Data deduplication is the technique to prevent the storage of redundant data in storage devices. Data deduplication is gaining much attention by the researchers because it is an efficient approach to data reduction. Deduplication identifies duplicate contents at chunk-level by using similarity-based comparison methods and eliminates redundant contents at chunk level. According to Microsoft research on deduplication, in their production primary and secondary memory is redundant about 50% and 85% of the data respectively and should be removed by the deduplication technology.

The rate of data growth is exceeding the decline of hardware costs. Database compression is one solution to this problem. For database storage, in addition to space saving, compression helps reduce the number of disks I/Os and improve performance, because queried data fits in fewer pages. [13] Research in this paper focuses mainly on reducing the storage usage of cloud which has been done by using data deduplication and minimizing replicas

without threatening the data reliability. Similarity Based Data Deduplication (SiDe) is presented to minimize replicas, in the Cloud. It is designed to increase the profit by the lowering the cost of cloud storage. The LZW algorithm is used for compression. Only two replicas will be stored on cloud storage with one replica being the original data and other will be the compressed version.

II. RELATED WORK

With all current approaches for data reliability, the most effective approach in the distributed storage system is data replication. Such existing approach has been proposed in [2], [3], [5]. Data replication schemes universally accepted in current commercial Cloud systems include Amazon S3 [1], HDFS [3], GFS [6]. Despite data replication being widely used, it has some disadvantage that it would require a large amount of storage resources resulting in increase of cost. Also, for long-term storage it could create even more than three replicas of the data, which limits its capability to lower storage consumption.

In [6], original data is stored in the block of same size (data partitioning) to improve the reliability and availability of data in the storage system. However, these researches considered a constant value of failure rate of storage devices rather than taking the changing disk failure rate patterns of storage devices.

Efforts for establishing data reliability have also been made in the software aspect. In [8], for minimization in the data loss rate of storage system analytical model of reliability in data and data replication schemes are proposed.

The disk reliability has been further studied for many years in both academia and industry [4], [9], [10]. The assumptions of many studies are based on the factor that failure rate of each disk is constant. For example, a current study that examines reliability in data with Markov chain models assumes that the failure rates of all disks in the storage system are the same [9]. Another known example of failure rate of disk is called a “bathtub” curve, wherein the early life of disk, failure rate is higher, falls during the first year, remains relatively constant for the remaining useful lifespan of the disk and rises again at the end of the disk’s lifetime [7], [10].

In [12], only one replica of the data is stored to reduce data storage consumption. This mechanism is called as PRCR approach. This approach provides similar or higher data reliability rate as compared to the conventional three replica strategy by periodically checking for replicas.

In [13] dbDedup mechanism has been used for online databases which use block level compression databases pages or oplog along with delta encoding. It has achieved 37x reduction in storage space and 61x when paired with block level compression but the disadvantage being does not give better results for large files.

In [15], SiLo is proposed, which uses similarity of data streams and locality of data streams to achieve throughput, elimination of higher duplicate rate and well-balanced load at extremely low RAM overhead. The purpose of SiLo is to eliminate similarity in data by grouping strongly associated small files into a segment, dividing large files and to dominate locality in the data streams. It distributes data to multiple backup nodes.

In [16] a digital trie is used to build and keep track of data traversed in a file for retrieving & showing the desired portion of the text for faster compression and decompression. It is able to achieve more compression than conventional LZW but the ratio is very small

III. ARCHITECTURE DIAGRAM

SiDe is a mechanism for managing reliability in data which can handle a large amount of data in the Cloud. SiDe is based on 1+2 replica scheme that indicates it stores original replica of the file for short duration and for the longer duration, it stores an original replica and compressed version of it. SiDe can be used for storing of cloud data with a minimum cost of storage, serving data reliability requirement with minimum replication technique. Fig.1 shows the architecture of SiDe and explained below:

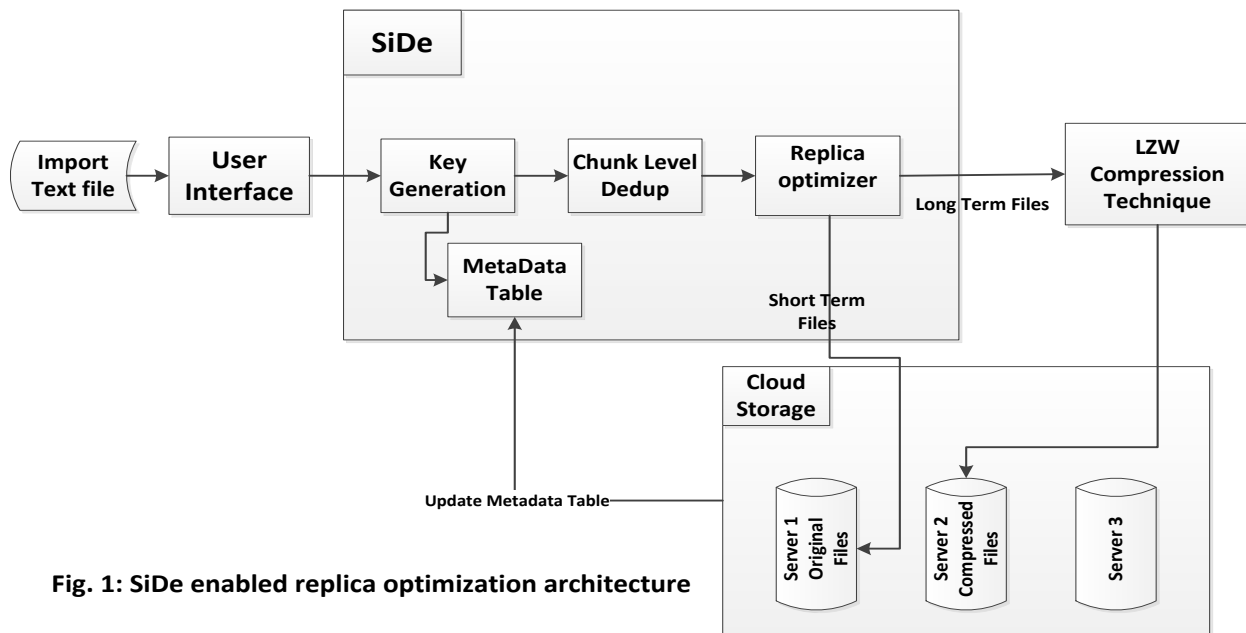


Fig. 1: SiDe enabled replica optimization architecture

A. User Interface

It is the component having various inputs from the user related to registration and login details apart from file operations such as upload files, view files, delete files etc. The storage duration of the file to categorize 1+2 replica scheme is taken from the user.

B. SiDe

The key components are Chunk level deduplication algorithm, Key Generation, Metadata Table and Replica Optimizer, LZW Compression and Decompression. These components are further explained in detail.

C. Chunk level deduplication algorithm

The chunk level deduplication algorithm is used to identify the duplicate parts within the files. In case of duplicate chunks being present, only one instance of the duplicated chunks is saved to avoid unnecessary storage space consumption by duplicate chunks. The regeneration code for each file is generated in order to recreate the original file when a download request is sent by the user.

Algorithm for chunking is as follows:

1. Divide the uploaded file into fixed sized chunks of 10 kB.
2. Assign unique id(character value) for each chunk;
3. Iterate over all the chunks and perform the similarity based comparison.
4. Duplicate chunks if found are assigned same id's.
5. Generate a regeneration code for the file in which the ids of chunks are present in the order in which the chunks appear in the file. Duplicate chunks have same ids whereas non-duplicate chunks have different ids.
6. In the cloud storage, store only the unique chunks and the regeneration code;

Regeneration algorithm is illustrated as follows:

1. Scan the regeneration code character by character.
2. At each character, use the chunk represented by that character to access the right chunk of the file stored in memory.
3. Merge all the chunks obtained.
4. Send the merged file to the user.

D. Key Generation

It is the component that collects metadata of files from the uploaded files and uses the metadata for generation of key. The generated key is used for the purpose of mapping the files uploaded by user on the server and the database where the metadata of the file is stored. Required parameters for key generation are file_name, client_id. Key Generation Algorithm is:

1. Calculate the ASCII value of file_name.
2. Fetch the 2nd character of file_name and append it with no. of characters in the string.
3. Append the result of step 1 and step 2 with client_id.

The obtained result will be the unique key generated for each file uploaded.

E. Metadata Table

For all files uploaded on the cloud storage, the metadata of the file is generated and stored in the metadata table. The attributes like file name, file size, username, regeneration code, etc. are stored in the table. All the file processing and management tasks are carried out by referring the metadata table.

F. Replica Optimizer

This component stores the replica in compressed and original version of data uploaded by the user on the cloud according to its storage period. Files with short term duration are stored in the form of one replica i.e. original file and files with long-term duration stored in the form of two replicas which is an original file uploaded and another is a compressed version of the original file.

G. LZW Compression and Decompression

When the new file is uploaded to the Cloud which is in a long-term storage category, that file is compressed using LZW compression technique creating a compressed version of the file. In case the original file gets deleted accidentally, it can be regenerated by decompressing the compressed version. This ensures a compressed copy is maintained as a backup for the original using minimal storage space.

Working of LZW algorithm is as follows:

1. An array of symbols is read and then the symbols are grouped into strings and convert them into object files which are binary.
2. As object files take less space than the strings, the strings get replaced, resulting in compression.
3. It uses an object file table. Single bytes from the input file are represented by the codes in the range from 0-255.
4. When encoding starts, the object file table consist of only the first 256 entries, with the remaining of the table being unused. Compression represents sequence of bytes by using object files from 256 to 4095.
5. In the further encoding, LZW analyses the repeated sequences in the data and inserts it into the object file table.
6. Decoding is done by taking each object file from the compressed file and converts with the help of object file table to find what character or characters it represents.

IV. PERFORMANCE ANALYSIS

A. Test Setup:

The proposed project is tested on 20-30 clients. The technology used for coding is JAVA. The IDE used for application development is NetBeans IDE 8.1. MySQL database has been used for maintaining the metadata of uploaded files. Text files are used as input.

B. Graphical Representation of Analysis:

1. Replica vs. data size

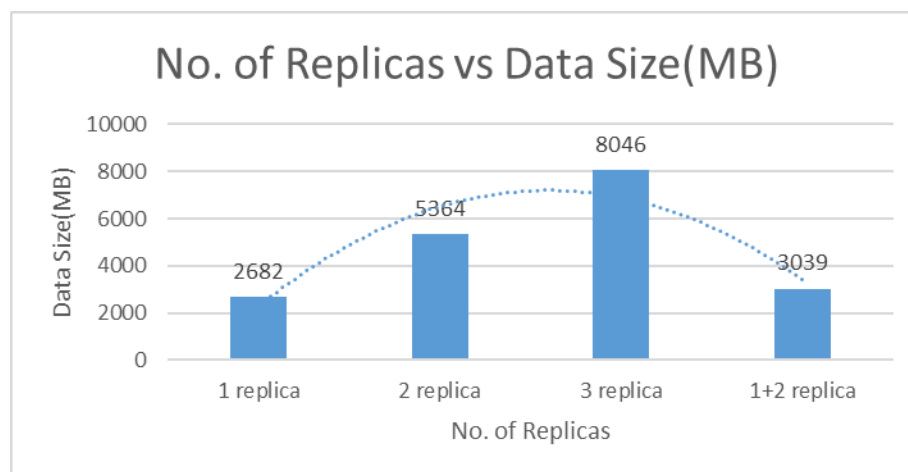


Fig.1 Replica vs. data size

The graph shown in fig.1 indicates how the number of replicas affects the storage space required for the data storage. It gives memory required for storing one replica i.e. Original file will be stored, two replica i.e. two copies

of the original file will be stored , three replica i.e. three copies of the original file will be stored , 1+2 replica i.e. one copy of the original file and one compressed file is stored. As the number of replicas increases, memory requirement goes on increasing. Those several replica schemes adopting different replication methods have been studied and compared with our implemented replication strategy i.e. 1+2 replica strategy and shows that our approach saves up to 62% storage space when compared with traditional 3 replica strategy.

2. Performance analysis of files without chunking algorithm

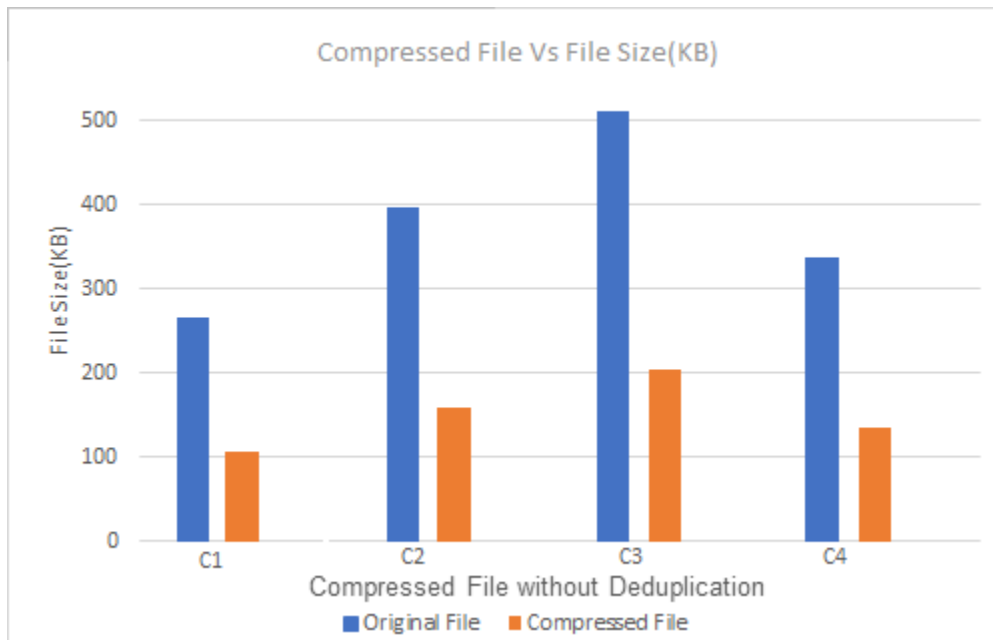


Fig. 2 Files without chunking algorithm vs. File Size

The fig.2 shows the graph of performance analysis of files without deduplication algorithm w.r.t clients i.e. C1 and so on. Results show that there is only 60-65% of reduction of the total storage which is due to the compression algorithm.

3. Performance analysis of files with chunking algorithm

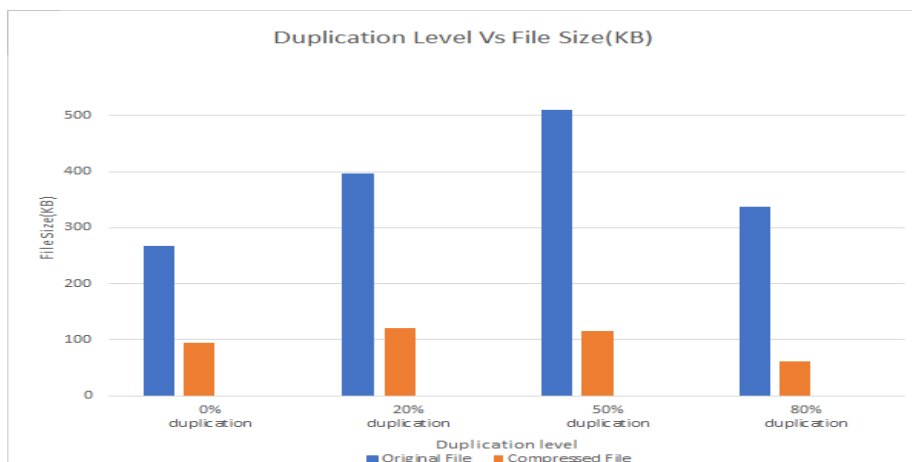


Fig 3. Files with chunking algorithm vs. File Size

The fig.3 illustrates the graph of 1+2 replica scheme where deduplication and compression are done using chunking and LZW algorithm respectively. The file having higher number of duplicate data chunks shows a total reduction of 73-79% and file having no duplicate data chunks show a total reduction of 65-70%, thus saving memory space by 81-84%.

4. Storage space on cloud vs. Clients

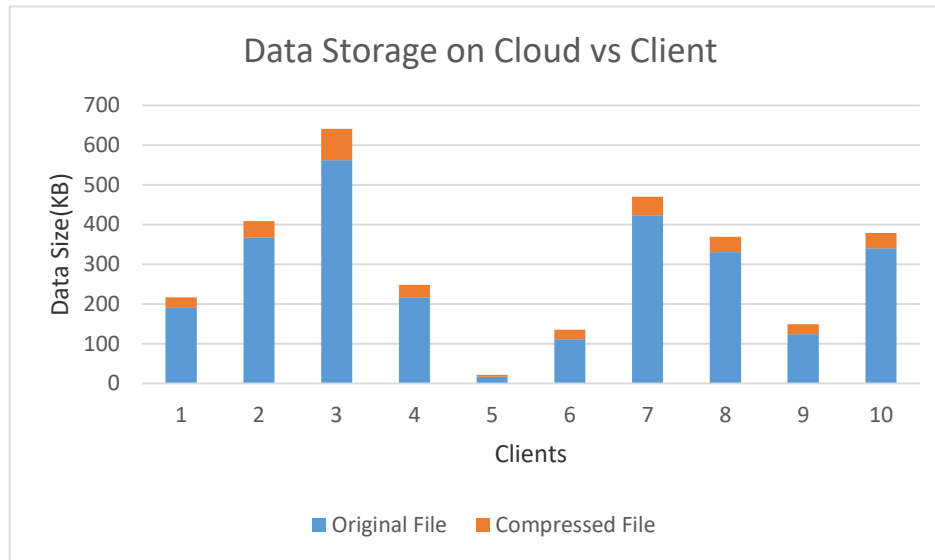


Fig.4 Data storage on cloud vs. Client

The graph fig.4 illustrates the how 1+2 replicas are stored on cloud according to each client files. The compressed files are stored in the cloud by going through the process of chunking and compression algorithm with client's original file.

V. CONCLUSION

In this paper, cost-efficient and data reliable management mechanism (SiDe) has been proposed and developed based on established data reliability model. It implements an interactive replica checking method to assure the reliability in data and maintains data with minimum two replicas (serving as a cost-efficient standard) one being the original replica and other being compressed version of the original replica based on the duration. Assessment of SiDe has proved that this mechanism is able to handle a vast amount of cloud data, considerably reducing 81% to 84% of the Cloud storage space for files varying size 10KB to 300KB and having duplicate data chunks. In future the SiDe can be extended based on the access frequency of file and further work can be carried out for inter-file level deduplication and for different types of files.

VI. REFERENCES

- [1] Amazon. (2011). Amazon Simple Storage Service (Amazon S3). [Online]. Available: <http://aws.amazon.com/s3/>, 2011.
- [2] B. Balasubramanian and V. Garg, "Fault tolerance in distributed systems using fused data structures," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 4, pp. 701–715, Apr. 2013.
- [3] D. Borthakur. (2007). The Hadoop Distributed File System: Architecture and Design [Online]. Available: http://hadoop.apache.org/common/docs/r0.18.3/hdfs_design.html.
- [4] J. G. Elerath and S. Shah, "Server class disk drives: How reliable are they?" in *Proc. Annu. Symp. Rel. Maintainability*, 2004, pp. 151–156.
- [5] A. Gharaibeh, S. Al-Kiswany, and M. Ripeanu, "ThriftStore: Finessing reliability trade-offs in replicated storage systems," *IEEE Trans. Parallel Distrib. Syst.*, vol. 22, no. 6, pp. 910–923, Jun. 2011.
- [6] S. Ghemawat, H. Gobioff, and S. Leung, "The Google file system," in *Proc. ACM Symp. Oper. Syst. Principles*, 2003, pp. 29–43.
- [7] G. Gibson, "Redundant disk arrays: Reliable, parallel secondary storage," Univ. California, Berkeley, CA, USA, Tech. Rep. UCB/ CSD 91/613, 1991.
- [8] M. Lei, S. V. Vrbisky, and Z. Qi, "Online grid replication optimizers to improve system reliability," in *Proc. IEEE Int. Parallel Distrib. Process. Symp.*, 2007, pp. 1–8.
- [9] S. Ramabhadran and J. Pasquale, "Analysis of long-running replicated systems," in *Proc. IEEE Conf. Comput. Commun.*, 2006, pp. 1–9.
- [10] Q. Xin, T. J. E. Schwarz, and E. L. Miller, "Disk infant mortality in large storage systems," in *Proc. IEEE Int. Symp. Model., Anal., Simul. Comput. Telecommun. Syst.*, 2005, pp. 125–134.
- [11] D. Yuan, Y. Yang, X. Liu, W. Li, L. Cui, M. Xu, and J. Chen, "A highly practical approach towards achieving minimum datasets storage cost in the cloud," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 6, pp. 1234–1244, Jun. 2013.
- [12] Wenhao Li, Yun Yang, Senior Member, IEEE, and Dong Yuan, Member, IEEE, "Ensuring Cloud Data Reliability with Minimum Replication by Proactive Replica Checking", *IEEE Transactions on Computers*, vol. 65, no. 5, May 2016.
- [13] "Online Deduplication for Databases", Lianghong Xu Andrew Pavlo Carnegie Mellon University Carnegie Mellon University lianghon@andrew.cmu.edu pavlo@cs.cmu.edu , Sudipta Sengupta Gregory R. Ganger, Microsoft Research Carnegie Mellon University sudipta@microsoft.com ganger@ece.cmu.edu.
- [14] James W. Anderson, Hein Meling, Alexander Rasmussen, Amin Vahdat, and Keith Marzullo, "Local Recovery for High Availability in Strongly Consistent Cloud Services", *IEEE Transactions on Dependable and secure computing*, VOL. 14, NO. 2, MARCH/APRIL 2017.
- [15] "Similarity and Locality Based Indexing for High Performance Data Deduplication" Wen Xia, Hong Jiang, Senior Member, IEEE, Dang Feng, Member, IEEE, and Yu Hua, Senior Member, IEEE.

[16] “Modified LZW Algorithm for Efficient Compressed Text Retrieval”- Nan Zhang, Tao Tao, Ravi Vijaya Satya, and Amar Mukherjee School of Computer Science, University of Central Florida {nzhang, ttao, rvijaya, amar}@cs.ucf.edu 2004 IEEE.

Wireless Automated Soil Monitoring System

Sadia Afrin Shampa

Dept. of CSE

Bangladesh University of Engineering & Technology
(BUET)

Dhaka, Bangladesh

sadia.cse.39@gmail.com

Buraira Nasir

.NET Developer,

Emerald Techno Limited

Dhaka, Bangladesh

khushboo.nasir@gmail.com

Umme Salma

Lecturer, Dept. of CSE

Bangladesh University

Dhaka, Bangladesh

ummesalma.cse@gmail.com

Abstract—Humidity is the amount of water present in soil and temperature is the amount of heat. They are inversely proportional to each other. The amount of water in soil is responsible for respiration, photosynthesis, transpiration and transportation of minerals and other nutrients through the plant. Proper irrigation is very important for soil for plant growth. On the other hand, it is also important to make a balance between the temperature and humidity. This thesis report presents an automated soil monitoring system where the sensor HSM-20G is being used as temperature and humidity sensor, low power HC-06 wireless transceiver an Atmel ARDUINO board implementing microcontroller ATMEGA-2560. The sensed soil contents from the HSM-20G sensor node will be transmitted to the coordinator via wireless signal. The coordinator will transmit the data to the mobile using RX interface. Main components are used in the proposed system are low cost and more flexible. The ATMEGA-2560 microcontroller consumes low power. In addition, the HSM-20G sensing electrode also has advantages of easy installation and replacement in the farm. The developed system provides a better data transmitted and processed wirelessly and it can serve as a basis for efficient irrigation scheduling and temperature control.

Keywords—Sensors, ZigBee, GPRS, GSM, GPS, Bluetooth, Arduino Board, WASMS

I. INTRODUCTION

Bangladesh is primarily an agricultural economy. Agriculture is the main producing sector of the economy, as it accounts for about 30% of the country's GDP and around 60% of the total labor force [1]. The performance of this sector has an overwhelming impact on the main macroeconomic objectives, such as job generation, poverty alleviation, human resource development and food security. We want to create a device that allows our farmers to obtain information about their land and other agricultural information in accordance with their various irrigation problems. For this purpose, we have developed a device "Wireless automatic floor monitoring system", so that each user only identifies the soil problems and obtains the solution. Only registered farmers receive this installation. Therefore, by registering a farmer is a member of this project and obtains that facility.

II. RELATED WORKS

In the last two decades, with the development of wireless technologies, several studies have focused on the autonomous irrigation with sensors in agricultural systems [10]. Among

these works, a micro sprinkler system has a different place, and is designed to block controlled solenoid valves in a citrus grove with wireless sensors. Many studies have also successfully demonstrated the use of remote active and passive microwave sensors. It has been discovered that many methods of irrigation planning have been developed by wireless sensors in the last decades. It has been discovered that many methods of irrigation planning have been developed by wireless sensors in the last decades. Many of the sensors, valves and modules available on the market assembled for irrigation system networks are too complex and / or expensive to be feasible for site-specific management of fixed irrigation systems. Its approval by producers is limited due to the costs, installation time, maintenance and complexity of the systems.

III. THEORETICAL CONSIDERATION

Temperature and humidity are two important elements of soil. Based on these two parameters we can find the condition of soil with respect to different crops. There are various types of sensors and these sensors have different interfacing circuits depending on the application. In wireless automated soil monitoring system we have use two circuits. One is determination of these parameters and another is Bluetooth device. This chapter reviews the theoretical consideration of our project.

A. Wireless (Bluetooth) Automated Soil Monitoring System

1) *Wireless Systems*: Wireless communication is a signal of transmitting data between two or more electric volcanoes. The best air conditioning for the air is Radio stations, portable, can be as short as several meters per television or even thousands or even millions of miles for wireless communication. It includes a variety of code, which can be accessed, which may include, including two phones, mobile phones, digital PDAs and wireless networks. Other

examples of radio broadcasts include GPS cells, computers, hardware hardware and audio help, door access, rocket, radio station, television, space and cell phones.

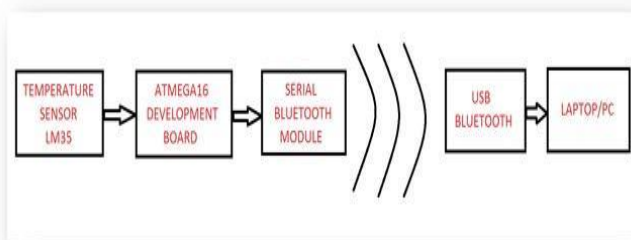


Figure 1. Example of a wireless connection

2) *System Automation*: It works in a variety of ways, including steel, hydraulic, pneumatic, electronics and computer systems. Advanced systems, such as modern airports and shipping companies, often combine all of these systems. Do not use a variety of methods, including metal, hydraulic, pneumatic, electronic and computer-based computers. Advanced systems, such as modern airports and shipping companies, often combine all of these systems.

B. Temperature and Relative Humidity Related Analysis

Humidity is the amount of water vapor that can be in the atmosphere at the given temperature. Relative humidity is the actual amount of water vapor in the atmosphere. As Humidity certain amount of water vapor in the atmosphere. As temperatures increase, the amount of water vapor that can be in the air also increases. As temperatures decrease, the amount of water vapor that can be in the air also decreases. In the winter time, the humidity can be so low it makes your skin itch because low humidity means dry air which also makes the barometer rise. The barometer measures the atmospheric pressure. As the atmospheric pressure rises, that would indicate clear weather and as it drops, that would indicate stormy weather [15].

Warm air can hold more moisture than cold air. Hence if the temperature of the air increases, its capacity to hold the moisture increases provided additional moisture is supplied to the air. That is why, in the tropic regions, the humidity (particularly absolute humidity) is constantly higher owing to the high temperature coupled with abundant moisture. Over the deserts, despite high temperature, humidity[9]. On the other hand, relative humidity (which is different from absolute humidity) decreases with a rise in temperature. Because the relative humidity depends not only upon the amount of water vapor actually present but also on the air temperature. Hence, if no moisture is added, an increase in temperature will result in a corresponding decrease in the relative humidity [9].

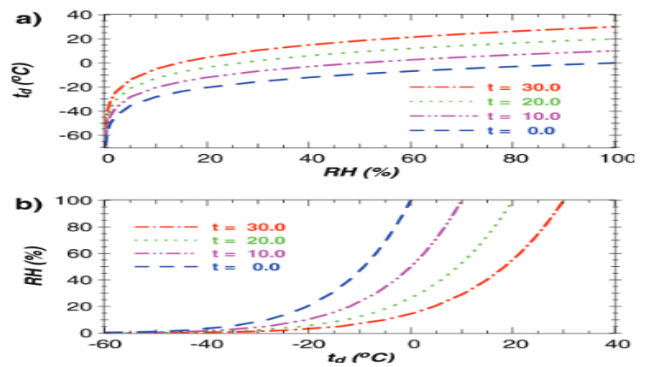


Figure 2. Relationship between dew point temperature and relative humidity

IV. METHODOLOGY

Figure 3 shows the block the diagram of the wireless automated soil monitoring system considering the Table 1. At first the device will be connected to the soil using probs. Then the temperature and humidity level of soil will be determined or sensed by the HSM-20G sensor. Humidity level has been shown in the form of percentage and the temperature is showed in both of the forms Celsius and Fahrenheit. The data of temperature and humidity is then sent to the mobile device through a wireless connection (in our project we have used HC-06 Bluetooth module). The Bluetooth device is a wireless media to pass information within components. Afterwards the information will be transferred to an android device. Then it will analyze the information and provide the corresponding result.

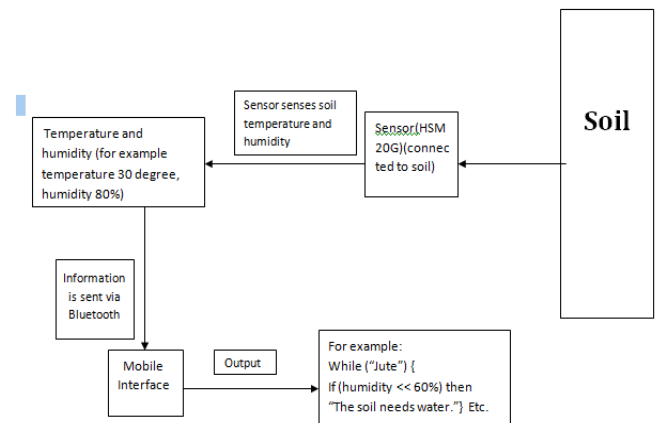


Figure 3. The Block diagram of WASMS model

A. System Design

The circuit diagram is very simple. The two Analog input pins A0 & A1 of Arduino board are used to measure the sensor output voltages that correspond to the ambient temperature and relative humidity.

1st portion is designed to determine the humidity and temperature from the soil with the HSM-20G: Humidity and Temperature Sensor.

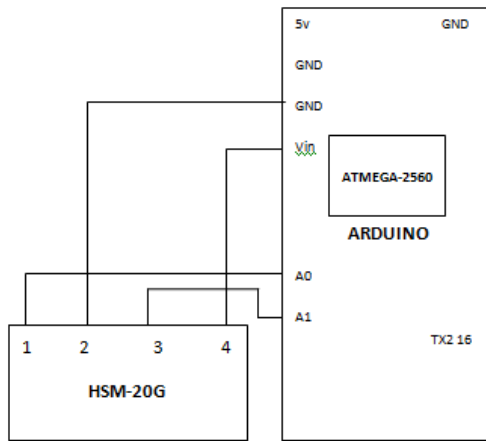


Figure 4. The Circuit Diagram of WASMS model (1st portion)

2nd portion is the establishment wireless connection between the sensor circuit and the android device via HC-06 Bluetooth device by interfacing it with the Arduino UNO.

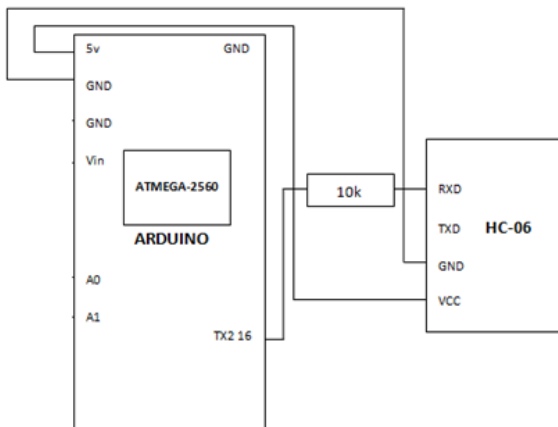


Figure 5. Interfacing Microcontroller to the Bluetooth device (HC-06)

Bluetooth is a popular way of communication between devices. Many smartphones now have the ability to communicate via Bluetooth. This is useful for developers of mobile applications whose applications require a wireless communication protocol.

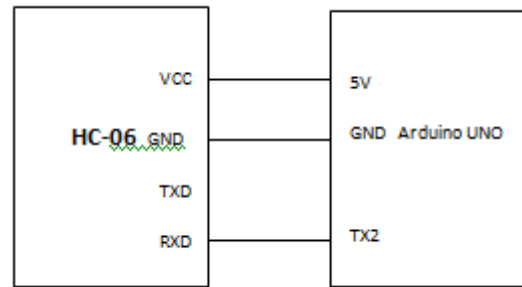


Figure 6. Bluetooth Module – Arduino Uno Connectivity

However, applying the Bluetooth API can be difficult for first-time users. The objective of this application note is to explain how to use the Bluetooth tools available to an Android developer in order to send and receive data to and from another device wirelessly.

B. System Architecture

1) Sensing soil parameter (Temperature & humidity) from soil:

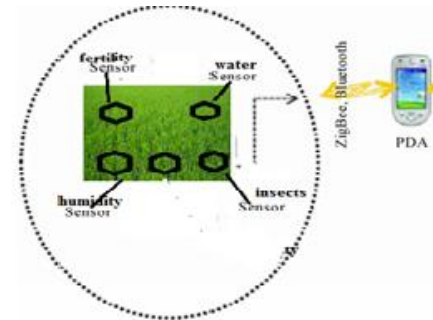


Figure 7. Architecture of WASMS

We have considered two parameters for determine the condition of the soil. And those are:

- Humidity determination from soil
- Temperature determination from soil

After sensing all parameters PDA (Personal Digital Assistant) gets all the information and stored it and sent that information to the internet via Bluetooth.

2) Wireless internet portion:

We are using a network for internet connection. Here listed (registered) farmer gets facility from the agriculture office. So it is necessary to sign up for the farmers.



Figure 8. Internet portion of Automated Soil Monitoring System

Temperature measurements can vary from a single instrument that registers the outside temperature in the shade to different measurements (for example on a standard screen, just above grassland or bare ground (soil temperature) [9], in a building and / or with a wet bulb, etc.).

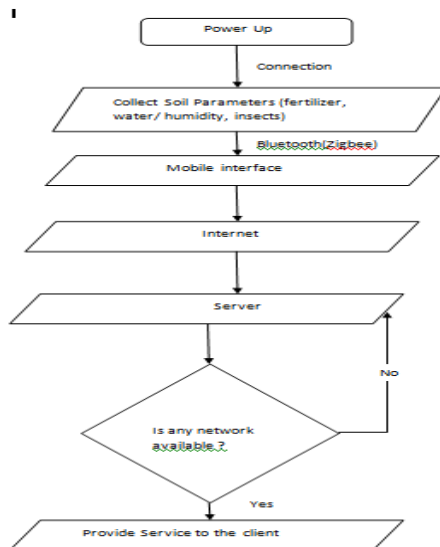


Figure 9. Flow chart of Wireless Automated Soil Monitoring System

C. Temperature and humidity measurement

We have collected some data on temperature and humidity for various crops. We have used the following table for required comparison:

| No. | Name of the Crops, Vegetables, Fruits, Flowers | Temperature (Celsius) | Temperature (Fahrenheit) | Humidity (%) |
|-----|--|-----------------------|--------------------------|--------------|
| 1. | Rice | 40 | 104 | 50-55< |
| 2. | Jute | 45 | 113 | 45-50< |
| 3. | Corn | 30 | 86 | 60-65< |
| 4. | Lemon | 50 | 122 | 85-90< |
| 5. | Tomato | 32 | 89.6 | 70-85< |
| 6. | Beans | 52 | 125.6 | 85-70< |
| 7. | Mango | 10-13 | 50-55.4 | 85-95< |
| 8. | Jackfruits | 25-30 | 77-86 | 85-95< |
| 9. | Pineapple | 10-13 | 50-55.4 | 85-95< |
| 10. | Coconut | 13-16 | 55.4-60.8 | 80-85< |
| 11. | Banana | 17-21 | 62.6-69.8 | 85-95< |
| 12. | Water lily | 50 | 122 | 50< |
| 13. | Rose | 26.66 | 79.98 | 70< |
| 14. | Tuberose | 18-32 | 64.98-89.6 | 22-83< |
| 15. | China rose | 7-40 | 44.6-104 | 20-80< |
| 16. | Marigold | 23.88-26.66 | 74.98-79.98 | 80-100< |

Table 1. Standard Table of temperature and humidity

In our project we have used following equations to calculate the humidity and temperature:

$$RH = 0.1515 * \text{sensorValue2Avg} - 12.0 \quad (1)$$

$$T_{inC} = 281.583 * \text{pow}(1.0230, (1.0/R)) * \text{pow}(R, -0.1227) - 150.6614 \quad (2)$$

$$T_{inF} = T_{inC} * (9.0/5.0) + 32 \quad (3)$$

Eq 1. is for humidity and eq2. and eq3. are for temperature in Celsius and Fahrenheit respectively.

D. Circuit Implementation

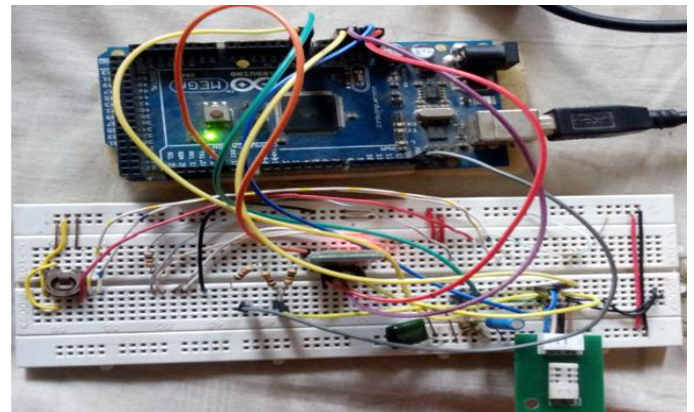


Figure 10. The circuit implementation of WASMS

This is the overall circuit implementation of our thesis project.

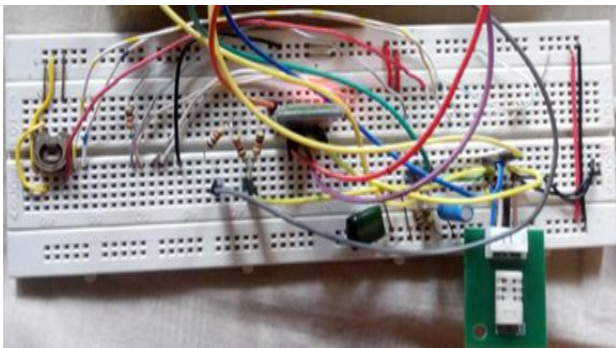


Figure 11. Partial circuit implementation with bluetooth module and sensor

This partial circuit shows the interfacing between temperature and humidity sensor (HSM-20G) and the bluetooth module (HC-06).

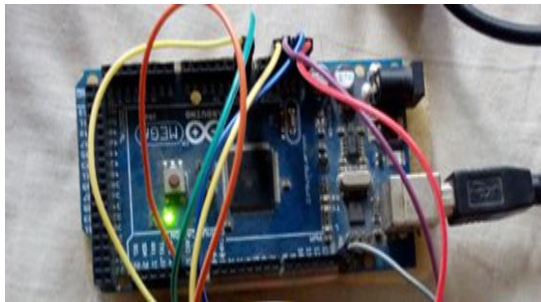
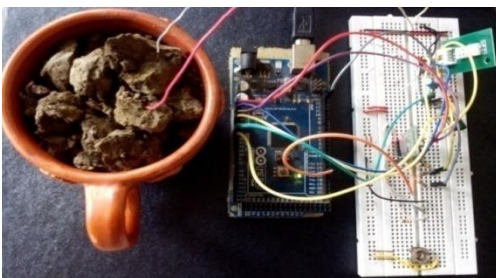


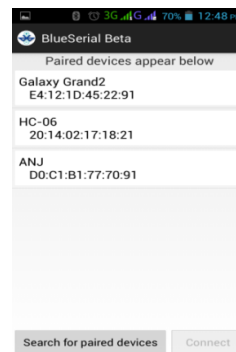
Figure 12. The arduino board

This section shows the implementation with microcontroller ATMEGA2560. By loading arduino program, this humidity and temperature sensor are being sent to the android device. At last the device gives a output by pairing android with bluetooth module.

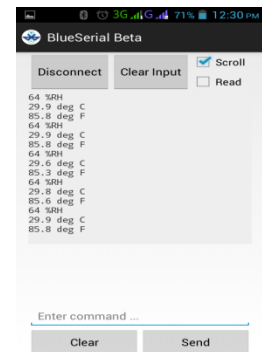
E. Result



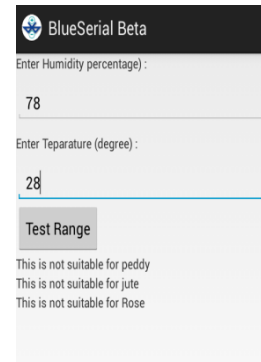
i)



ii)



iii)



iv)

Fig 13. i) Main Circuit, ii) Pairing bluetooth module and android device, iii) Received data from soil, iv) Corresponding Result

V. DISCUSSION AND RECOMMENDATION

A. Discussion

The main objective of this research was to merge the idea of wireless technology and system automation for the purpose of monitoring the soil parameters such as humidity and temperature.

The publications on this idea was being studied. The process of soil monitoring system is studied. The idea of wireless system and automation has been used with the conventional SMS.

Bluetooth module (HC-06) is used for wireless transmission and Temperature and humidity sensor (HSM-20g) is used as the main sensor. An android program is developed to receive the data and show the result expected.

Block diagram of the system was designed. Simplified circuit of the functional components of the system was also designed and simulated.

B. Future Recommendation

The humidity and temperature vary from soil to soil[8]. Also different crops required different humidity and temperature. The system can be developed for every variations. A central server can provide informations more fluently. So our recommendations for future are:

- Collect the standard values of all possible soil parameters and create a standard database.
- Extending the system for all types of soil and crops.
- Use an online server to provide the information through the internet with the help of mobile operators.

Develop a system suitable for all types of mobile OS such as Java, Symbian etc.

REFERENCES

- [1] Economy_development_in_Bangladesh, <http://www.assignmentpoint.com/business/economics/economy-development-in-bangladesh.html>.
- [2] <http://agrobangladesh.com/>.
- [3] http://www.theglobaleconomy.com/Bangladesh/Agriculture_productivity/.
- [4] B.T. Kang, B. Tripathi, Technical paper 1: Soil classification and characterization, ILRI.
- [5] Poverty in Bangladesh, http://en.wikipedia.org/wiki/Poverty_in_Bangladesh.
- [6] Mohammadh. Mondal, crop agriculture of Bangladesh: challenges and opportunities, ISSN 0258-7122, Bangladesh J. Agril. Res. 35(2): 235-245, June 2010.
- [7] Shamsul Alam, State of the economy by June 2013: An appraisal-I, Published: Wednesday, 14 August 2013.
- [8] Humidity, <http://en.wikipedia.org/wiki/Humidity>.
- [9] Temperature, <http://en.wikipedia.org/wiki/Temperature>.
- [10] Aniket H. Hade, Dr. M.K. Sengupta, Automatic Control of Drip Irrigation System & Monitoring of Soil by Wireless, IOSR Journal of Agriculture and Veterinary Science (IOSR-JAVS).
- [11] Md. Imam Hossain, Climate Change: A Challenge to Coastal Agriculture in Bangladesh.
- [12] Wen-Yaw Chung, Jocelyn F. Villaverde, Janine Tan, Wireless Sensor Network Based Soil Moisture Monitoring System design, National Science Council, Taiwan (NSC 102-2221-E-033-066).
- [13] MARK G. L. AWRENCE, the Relationship between Relative Humidity and the Dew Point Temperature in Moist Air a Simple Conversion and Application.
- [14] Wireless, <http://en.wikipedia.org/wiki/Wireless>.
- [15] <http://www.engineersgarage.com/contribution/expert/wireless-temperature-monitoring-system-using-bluetooth>.
- [16] Rifkin, Jeremy (1995). The End of Work: The Decline of the Global Labor Force and the Dawn of the Post-Market Era. Putnam Publishing Group. pp. 66, 75. ISBN 0-87477-779-8.
- [17] Bennett, S. (1993). A History of Control Engineering 1930-1955. London: Peter Peregrinus Ltd. On behalf of the Institution of Electrical Engineers. ISBN 0-86341-280-7.
- [18] Automation, Wikipedia, <http://en.wikipedia.org/wiki/Automation>.
- [19] Arduino, <http://en.wikipedia.org/wiki/Arduino>.

Cloudlet based Healthcare and Medical Knowledge Extraction System for Medical Big Data

Dr. Anjali A. Chandavale
Department of Information Technology
MIT College of Engineering
Pune, India
c.anjali38@gmail.com

Anuja J. Gade
Department of Information Technology
MIT College of Engineering
Pune, India
anujagade08@gmail.com

Abstract—Along with improvement of clouds, cloudlet and wearable devices, it becomes necessary to provide better medical information sharing over the internet. As we know, sharing medical information is very critical and challenging issue because medical information contains patient's delicate information. The medical information sharing mainly includes collection, storage and sharing of medical information over the internet. In existing healthcare framework, patient asks health query/question which is being answered by multiple doctors. The user is provided with correct answer with the help of truth discovery method. The medical information and history of patient is directly delivered to the remote cloud. The paper proposes system to provide confidentiality of medical information and protect healthcare system from intrusion. In this proposed system, during the medical information collection stage, the Number Theory Research Unit (NTRU) algorithm is used to encrypt user's vital signs collected various sensors. Then to protect medical information stored at remote cloud against the malicious attacks, a Collaborative Intrusion Detection System (CIDS) built on cloudlet mesh is proposed. Medical Knowledge Extraction (MKE) System is proposed to provide most reliable answer to the user's health related query. MKE extract the medical knowledge from noisy query-answers pairs and estimate the trustworthiness degree and doctor expertise using truth discovery method.

Keywords- Healthcare, Confidentiality, Intrusion, Number Theory Research Unit (NTRU), Collaborative Intrusion Detection System (CIDS), Medical Knowledge Extraction (MKE), Truth discovery.

I. INTRODUCTION

With the growth in cloud computing [2], as well as healthcare big data and wearable technology [1], cloud-based healthcare big data computing becomes difficult to provide user's health discussion associated demands [9] [10]. With the advances in clouds, cloudlet technology and wearable technology, it is essential to deliver well medical information sharing over the internet. As we know sharing this medical information on social network is beneficial for patients as well as doctors, the patient's delicate information might be disclosed or taken which results in privacy and security problems [11]

[12] without providing confidentiality to the shared information. Therefore it becomes challenging to balance confidentiality of medical information along effective medical information sharing.

In existing healthcare framework, the medical information which involves user's delicate information was delivered to the remote cloud which causes communication energy consumption. Further cloud-based medical information sharing addresses the following problems:

1. How to provide confidentiality to the user's body information collected by various sensors during its transmission to a nearby cloudlet?
2. How to offer security to the healthcare information stored in a remote cloud?
3. How to defend the healthcare framework from intrusion?

With the fast development in technology, today's young generation always prefer to search health related information, doctor's suggestion on any health related problems through the internet. Today, large numbers of health associated queries are searched over internet each day. Number of patients and doctors are involved in the medical crowd sourced query answering website in recent years. The noisy query-answers pair and filtering out unrelated or incorrect information are major challenges while extracting the medical knowledge.

Cloudlet based healthcare and medical knowledge extraction system is proposed to overcome the existing system issues, such as to reduce the communication energy consumption, to protect the whole healthcare system from intrusion and to provide most reliable answers to the patient. In this proposed system, NTRU algorithm [6] is utilized for encrypting the user's body information to provide confidentiality. Collaborative IDS [4] [5] built on cloudlet mesh is presented to defend whole healthcare system from intrusion. To provide most reliable answer to the user, MKE system [8] is proposed. Using truth discovery method, MKE system provides high quality knowledge triples and estimates the doctor expertise.

A cloudlet is small scale cloud data center to quickly provide cloud computing resources to the mobile devices such as smartphones, tablets, wearable device. Cloudlet represents a middle tier of three hierarchies: Mobile device-cloudlet-cloud [14]. Figure 1 shows the Mobile device-Cloudlet-Cloud architecture.

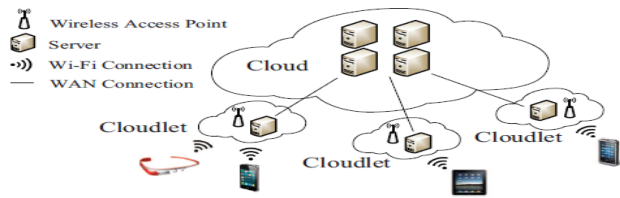


Figure 1 Mobile device-Cloudlet-Cloud architecture

Cloudlets have connectivity with Internet to provide resources to the close mobile devices. Cloudlet can be view as a data center to bring cloud closer. The goal of the cloudlet is to increase response time of applications which are latency delicate and resource demanding such as face recognition, augmented reality by hosting cloud computing resources such as virtual machine physically closer to the mobile device by using the Wireless Local Area Network (WLAN) [15].

Cloudlet is a type of edge computing. The goal of the edge computing is to offer the distributed computing and storage capability to the mobile devices at the network edge. Fog computing and Mobile edge computing are the another types of edge computing implementations. Comparison of cloudlet with fog computing and mobile edge computing is presented in [15]. Table 1 shows the comparison.

Table 1
Comparison between Cloudlet, Fog Computing and Mobile Edge Computing

| Parameters | Cloudlet Computing | Fog Computing | Mobile-Edge Computing |
|-------------------------|----------------------|-----------------------------------|-----------------------|
| Node devices | Act as a data center | Routers, Switches | Servers |
| Context awareness | Low | Medium | High |
| Proximity | One Hop | One or Multiple Hops | One Hop |
| Access Mechanisms | Wi-Fi | Bluetooth, Wi-Fi, Mobile Networks | Mobile Networks |
| Internode Communication | Partial | Supported | Partial |

Mobile Cloud Computing is the grouping of cloud computing and mobile computing by using the wireless network

connectivity in order to bring the computational resources to the mobile user. Mobile Cloud Computing overcome the many challenges such as low computing power, limited storage and short battery life by providing the computation offloading and data offloading from mobile device to the cloud. Again mobile cloud computing faces many challenges such as restricted bandwidth, cost and latency. In order to overcome these challenges, mobile cloud computing based on cloudlet technology comes into picture.

Cloudlet based computation offloading is a technique to supplement the computing capacity of mobile devices by the migration of computation to the external computing platform i.e. cloudlet [14]. Cloudlet based computation offloading is useful for applications which are latency delicate and resource demanding such as face recognition and augmented reality. Computation offloading is also known as cyber foraging.

Cloudlet based data offloading is a technique for improving data transfer between the mobile device and cloud. This is achieved by caching the data in cloudlet [14]. This technique is useful for data demanding applications such as video on demand, video surveillance and cloud storage. Drop box is an example cloud storage application.

Aymen El Amraoui [16], proposed Cloudlet softwarization architecture for pervasive healthcare system. One of the most important applications of sensors network is patient monitoring. Wireless Body Area Networks (WBANs) plays vital role in patient monitoring. WBAN is a collection of different wireless sensors that are implanted on or in patient body for monitoring the different physiological conditions of patients such as blood pressure, temperature, pulse rate etc. But in some of the critical situation of the patient, it is important to take an immediate action to save the patient life. So in such situation, it is important to provide fast and effective healthcare service. Therefore author proposed the new architecture based on the combination of cloudlet and WBANs. In this architecture, patient data is extracted through the cloudlet, so that user can access e-healthcare services at competitive cost. Face recognition application is based on the cloudlet based computation offloading because it is latency delicate and resource demanding mobile application. Therefore face recognition application using the cloudlet technology is presented in [17], which overcomes the limited resource availability problem of mobile device. In order to handle the big amount of data created by Body Area Network which is the collection of different sensors, author proposed a new framework based on cloudlet technology [3]. The proposed framework provide accessible storage and processing capability by means of middle tier i.e. cloudlet. So in this paper, we present one of the applications of cloudlet technology as a cloudlet based healthcare and medical knowledge extraction system for medical big data in healthcare domain.

The remaining of this paper is structured as follows. Section 2 presents the related work whereas in section 3 present

the proposed system based on the cloudlet technology. Further next section 4, mainly focus on the implementation of the proposed system. The experimental results and performance analysis is presented in section 5. Final conclusions are presented in last section 6.

II. RELATED WORK

In traditional cloud-based healthcare framework, big data computing became critical and complex to meet user's growing demands on health discussion. As we know, sharing medical information on social network is useful for patients as well doctors but sharing of the patient's medical information might be taken or disclosed which results in privacy and security problems.

Cloud – based medical information sharing address the problems such as: privacy protection of user's body data, providing security to the healthcare big data stored in remote cloud and detection and prevention of malicious attacks to protect the whole healthcare system.

Following papers are referred to overcome these challenges.

K. Hung [1], proposed a tele-home healthcare system which uses wearable devices to track the health related information i.e. physiological conditions of user, multi-sensor data fusion methods and wireless communication technologies.

M. S. Hossain [2], proposed a cloud-supported cyber-physical localization system. The goal of the proposed system is patient monitoring. The patient monitoring is done using smart-phones to track the voice and electroencephalogram signs of user in accessible, real-time, and effective way.

M. Quwaider [3], proposed cloudlet-based effective data collection framework in a wireless body area networks. The proposed framework tries to reduce the end-to-end packet interruption by selecting dynamically a nearby cloudlet, so that

the total interruption is reduced. Advanced CloudSim simulator is used to estimate the performance of framework.

H. Mohamed [4], proposed a collaborative intrusion detection and prevention system to identify and block various types of attacks and intrusions in order to protect the whole healthcare system.

Y. Shi [5], for protecting mobile clouds from intrusion and networks attack and securing the infrastructures among mobile devices, cloudlet, and remote clouds, proposed a series of authentication, authorization, and encryption protocols.

K. Rohloff [6], presented a fully homomorphic encryption scheme based on the Number Theory Research Unit (NTRU) algorithm. The presented scheme minimizes the frequency of bootstrapping operations.

Min Chen [7], proposed a cloudlet based healthcare system in order to defend the security of user's body data collected by wearable devices, to protect the healthcare big data stored in remote cloud and to successfully defend whole healthcare system from malicious attacks.

Yaliang Li [8], proposed a Medical Knowledge Extraction (MKE) System to deliver high quality knowledge triples and to estimate doctor expertise using truth discovery method.

III. PROPOSED SYSTEM

The basic idea behind the proposed cloudlet based healthcare and medical knowledge extraction system is to provide confidentiality and intrusion avoidance for cloudlet based medical information sharing over internet. This system also provides most reliable answers to the patient's health related query. The system is built up by utilizing the flexibility of cloudlet.

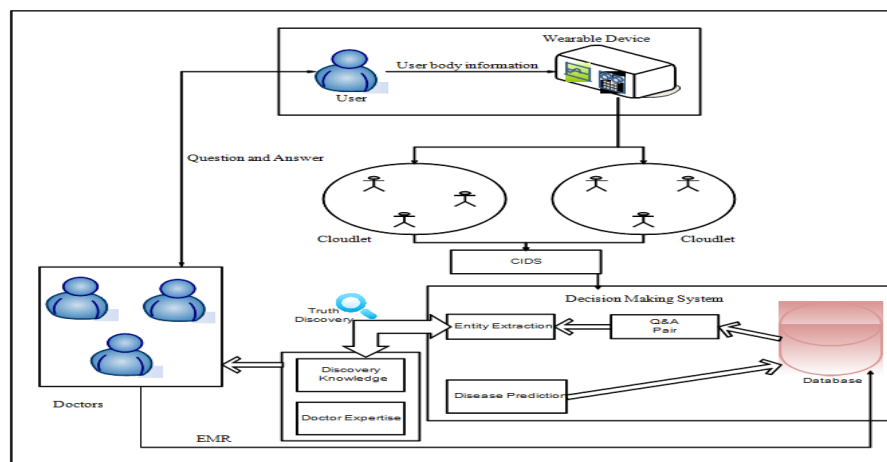


Figure 2 Proposed system architecture

As shown in figure 2 the proposed system works as follows:

Data collection by wearable device: The body information i.e. physiological conditions of user collected by wearable devices i.e. by various sensors are protected by NTRU algorithm in order to provide protection before it is transmitted to the cloudlet. The encrypted data will then store in nearby cloudlet through cellular network or Wi-Fi. In the proposed system, we track the blood pressure, pulse rate and temperature of the user through the sensors.

Collaborative Intrusion Detection System (CIDS): CIDS is designed among N different Intrusion Detection System (IDS) in order to get high intrusion detection rate. The N IDS are supposed to detect independently. Before transmitting medical information to remote cloud, CIDS complete the intrusion detection task. Once a malicious attack is identified, CIDS will fire an alarm or block the visit.

Medical information privacy protection in remote cloud: Remote cloud contains information of patients treated in hospital. Data or information stored in remote cloud is in Electronic Medical Records (EMR) form. In this stage EMR are divided into 3 classes:

1. EID (Explicit Identifier) - EID is the property which can identify users apparently such as name, Phone no, Email, Home Address.
2. QID (Quasi-Identifier) – QID is the property which can identify users approximately such as date of birth, gender.
3. MI (Medical Information) – Medical information such as disease type and disease symptoms.

To protect privacy of medical data stored in form of EMR at remote cloud, EID and QID are encrypted by using NTRU method.

Disease prediction: As doctor has access to remote cloud, a disease prediction is done based on the ranges of user's physiological conditions. The prediction will report to users.

Medical Knowledge Extraction System (MKE): MKE extract knowledge triples <query, diagnosis, trustworthiness degree> from Q-A pair using truth discovery method. To apply truth discovery method, first entities are extracted from query and answers texts and transform it into the entity based representation <query, diagnosis, source>. The aim of truth discovery is to resolve the conflicts and find truth i.e. most reliable answer for each query by estimating doctor expertise.

IV. IMPLEMENTATION

For the proposed system, we have used two algorithms, first is Number Theory Research Unit algorithm which is used to encrypt the user's body information for privacy protection and second is MKE system algorithm for the provision of trustworthy answers to the patients.

Algorithm 1: Number Theory Research Unit

Input: u, v, Message (m).

Output: encrypted and decrypted message.

Step 1: Two small polynomial u and v.

Step 2: The large modulo j and modulo k.

Step 3: The inverse of u modulo k and the inverse of u modulo j.

Step 4: $u * uk = 1 \pmod{k}$ and $u * uj = 1 \pmod{j}$

Step 5: Creating $uj = u^{-1} \pmod{j}$ and $uk = u^{-1} \pmod{k}$.

Step 6: Using j, uk and v, calculate the private key pair and the public key h.

Step 7: $h = juk * v \pmod{k}$.

Step 8: Encrypted message e is created using m, r and h as follows: $e = r * h + m \pmod{k}$.

Step 9: The private key u is used to calculate: $x = u * e \pmod{k}$.

Step 10: $z = uj * y \pmod{j}$

The polynomial z will be equal to the original message, if decryption procedure has been successfully finished.

NTRU Key generation:

The private and public key pair is created using the NTRU key generation scheme. The key generation method starts by selecting two small polynomials u and v, where small is well-defined as having coefficients smaller than the large modulo j and modulo k. The user must calculate the inverse of u modulo k and the inverse of u modulo j such that $u * uk = 1 \pmod{k}$ and $u * uj = 1 \pmod{j}$. The inverse of u is calculated both modulo j and modulo k, creating $uj = u^{-1} \pmod{j}$ and $uk = u^{-1} \pmod{k}$. The values of u and uj are taken as the private key pair and the public key h is calculated. The public key is calculated as follows:

$$h = juk * v \pmod{k} \dots \dots \dots (1)$$

NTRU Encryption:

The encryption procedure begins by creating a polynomial message m whose coefficients lie in an interval of length k. A small polynomial, r, is then created and used to obscure the message. The final encryption uses m, r and the public key h to create encrypted message e as follows:

$$e = r * h + m \pmod{k} \dots \dots \dots (2)$$

NTRU Decryption:

The decryption procedure uses the private key u to calculate:

$$x = u * e \pmod{k} \dots \dots \dots (3)$$

The coefficients of x must be selected in appropriate interval of length k to guarantee the highest probability that the decryption procedure will be successful. Once the coefficients of x are selected on the appropriate interval, x is reduced modulo j and the second private key is used to calculate:

$$y = x \pmod{j} \dots \dots \dots (4)$$

$$z = uj * y \pmod{j} \dots \dots \dots (5)$$

The polynomial z will be equal to the original message, if decryption procedure has been successfully finished.

Algorithm 2: Medical Knowledge Extraction System

Input: set of health related queries Q and their equivalent answers $\{x_q^d\} q \in Q, d \in D$, an external entity dictionary with

entity types, and real- value vector representations of entities.

Output: discovered knowledge triples $\langle \text{query}, \text{diagnosis}, \text{trustworthiness degree} \rangle$, and doctors expertise $\{w_d\}$.

Step 1: Segmentation: Extract words/entity from string;

Step 2: Entity extraction: extract similar meaning words (for example, illness) from query asked by patient and another type of entity (for example, disease) from the reply of doctor.

Step 3: input tuple creation for truth discovery method: form tuples $\langle \{ \text{entities from query text} \}, \text{entity from answer text}, \text{doctor ID} \rangle$

Step 4: set doctors' expertise uniformly;

Step 5: Compute the trustworthiness degree of each answer.

Step 6: Estimate doctor expertise w_d

Step 7: Discovered knowledge triples $\langle \text{query}, \text{diagnosis}, \text{trustworthiness degree} \rangle$ and the estimated doctor expertise $\{w_d\}$.

V. RESULTS AND PERFORMANCE ANALYSIS

Performance of NTRU algorithm

Number Theory Research Unit algorithm is used implement the public-key cryptography. It is computationally fast and

efficient method of data encryption. It lets faster encryption and decryption and simple implementation.

In proposed system, encryption of user's body information (physiological conditions such as temperature, blood pressure and pulse rate) which is collected by wearable device is done using the NTRU algorithm. NTRU algorithm is also used to perform data encryption at remote cloud. We evaluated the performance of the NTRU algorithm. We compared the changes in delivery ratio of NTRU algorithm and RSA algorithm with increase of time. Figure 3 shows the performance of NTRU algorithm.

Table 2
Performance of NTRU algorithm and RSA algorithm in terms of delivery ratio with increase of time

| Time(min) | Remote cloud's Delivery Ratio | User end's Delivery Ratio | RSA algorithm's delivery ratio |
|-----------|-------------------------------|---------------------------|--------------------------------|
| 0 | 0 | 0 | 0 |
| 1 | 0.5 | 0.3 | 0.2 |
| 2 | 0.73 | 0.7 | 0.65 |
| 3 | 0.83 | 0.8 | 0.76 |
| 4 | 0.84 | 0.81 | 0.79 |
| 5 | 0.86 | 0.83 | 0.8 |
| 6 | 0.93 | 0.9 | 0.86 |
| 7 | 1 | 1 | 0.89 |

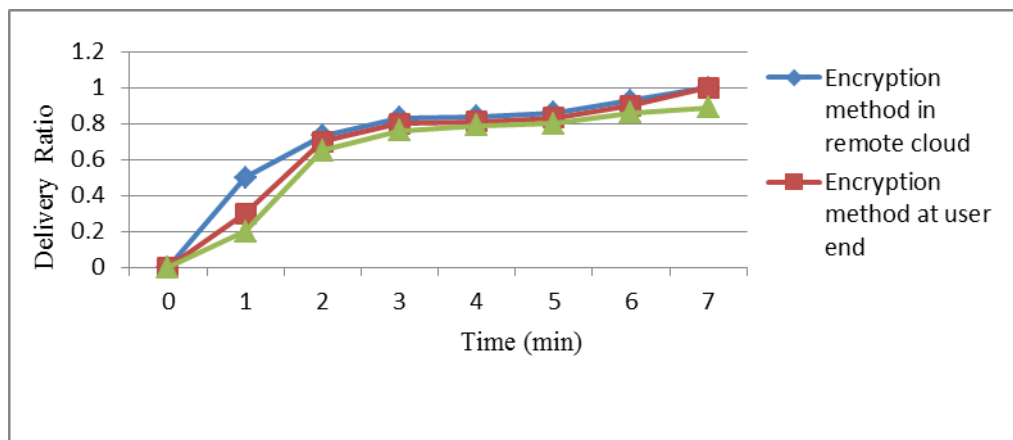


Figure 3 Comparison of the delivery ratio of NTRU algorithm with RSA algorithm

From the figure 3 and table 2, we conclude that NTRU algorithm offer better performance in terms of delivery ratio than delivery ratio of RSA algorithm.

Evaluation of MKE System

In the Medical Knowledge Extraction system, we have used QA.json dataset crawled from the website <https://answers.webmd.com/ask>. Data set is used for the comparison of various entities extracted from query asked by the user and the answers provided by different doctors to that query. Extracted entities of user and doctors are compared with entities in Q-A pairs of dataset. If the extracted entities matched with Q-A pairs, based on that the reliable answer is provided to the user. The reliable answer belongs to the corresponding doctor; the trustworthiness degree of that particular doctor is incremented. According to trustworthiness degree, doctor expertise is updated.

In MKE System, truth discovery method is used to find most reliable answer to the query asked by the user. First, different entities such as query, diagnosis and source are extracted from the query asked by the user and the answers provided by the different doctors to that query. In this system, source is nothing but the doctor's ID who gave answer to the query asked by the user. On that extracted entities, truth discovery method is applied to obtain the knowledge triples such as query, diagnosis and trustworthiness degree. Trustworthiness degree is calculated using following equation:

$$T(x_q) = \sum_{d \in D} w_d \cdot 1(x_{q,x_q^d}) \quad \dots\dots\dots(6)$$

Based on trustworthiness degree, doctor expertise is estimated using the following equation:

$$w_d = -\log \left(1 - \frac{\sum_{x \in V_d} T(x)}{|V_d| + c_{pseudo}} \right) \quad \dots\dots\dots(7)$$

From figure 4, we conclude that whenever users ask any health related query, the trustworthiness degree is calculated for each answer provided by doctors. Based on that, doctor expertise is estimated an updated. From the table 3, we conclude that the Doctor 2 has the highest expertise than the other doctors.

Table 3
Estimated Doctor Expertise based on trustworthiness degree

| Doctor ID | Estimated Doctor Expertise |
|-----------|----------------------------|
| Doctor 1 | 0.2 |
| Doctor 2 | 0.7 |
| Doctor 3 | 0.5 |

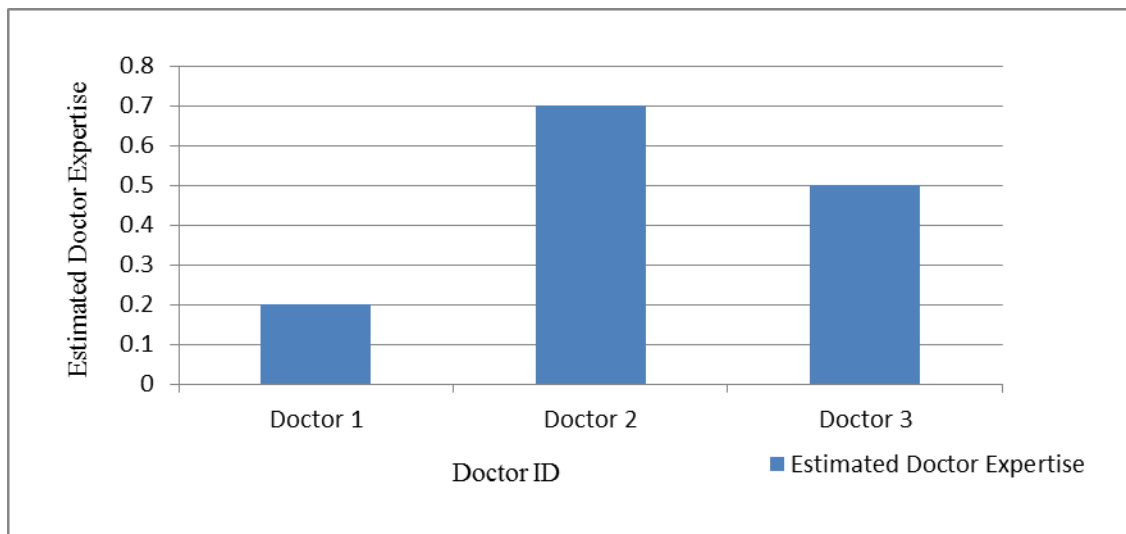


Figure 4 Comparison of estimated doctor expertise using MKE system

VI. CONCLUSION

In this paper, proposed and developed cloudlet based healthcare and medical knowledge extraction system provide confidentiality to the user medical information using NTRU algorithm. Proposed algorithm provides better performance in terms of delivery ratio than RSA algorithm. In MKE system, using the truth discovery method, most reliable answer is provided to the user's query by calculating the trustworthiness degree of answer. Based on that trustworthiness degree, each and every time, doctor expertise is calculated and updated. Finally in order to protect the whole healthcare system from intrusion, collaborative intrusion detection system is used.

REFERENCES

- [1] K. Hung, Y. Zhang, and B. Tai, "Wearable medical devices for tele-home healthcare", in *Engineering in Medicine and Biology Society, IEMBS'04.26th Annual International Conference of the IEEE*, vol. 2, IEEE, pp. 5384–5387, 2004.
- [2] M. S. Hossain, "Cloud-supported cyber-physical localization framework for patients monitoring", 2015.
- [3] M. Quwaider and Y. Jararweh, "Cloudlet-based effective data collection in wireless body area networks", *Simulation Modelling Practice and Theory*, vol. 50, pp. 57–71, 2015.
- [4] H. Mohamed, L. Adil, T. Saida, and M. Hicham, "A collaborative intrusion detection and prevention system in cloud computing", in *AFRICON*, IEEE, pp. 1–5, 2013.
- [5] Y. Shi, S. Abhilash, and K. Hwang, "Cloudlet mesh for securing mobile clouds from intrusions and network attacks", in *The Third IEEE International Conference on Mobile Cloud Computing, Services, and Engineering*, 2015.
- [6] K. Rohloff and D. B. Cousins, "A accessible implementation of fully homomorphic encryption built on ntru", in *Financial Cryptography and Data Security*. Springer, pp. 221–234, 2014.
- [7] Min Chen; Yongfeng Qian; Jing Chen; Kai Hwang; Shiwen Mao; Long Hu, "Privacy Protection and Intrusion Avoidance for Cloudlet-based Medical Data Sharing", *IEEE Transactions on Cloud Computing*, Volume: PP, Issue: 99 Pages: 1–1, 2017.
- [8] Yaliang Li; Chaochun Liu; Nan Du; Wei Fan; Qi Li; Jing Gao; Chenwei Zhang; Hao Wu, "Extracting Medical Knowledge from Crowdsourced Query Answering Website", *IEEE Transactions on Big Data*, Volume: PP, Issue: 99 Pages: 1–1, 2016.
- [9] R. Zhang and L. Liu, "Security models and requirements for healthcare application clouds", in *Cloud Computing (CLOUD)*, *IEEE 3rd International Conference on*, IEEE, pp. 268–275, 2010.
- [10] J. Zhao, L. Wang, J. Tao, J. Chen, W. Sun, R. Ranjan, J. Kołodziej, A. Streit, and D. Georgakopoulos, "A security framework in g-hadoop for big data computing across distributed cloud data centres", *Journal of Calculator and System Sciences*, vol. 80, no. 5, pp. 994–1007, 2014.
- [11] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data", *Parallel and Distributed Systems, IEEE Transactions on*, vol. 25, no. 1, pp. 222–233, 2014.
- [12] C. Zhang, J. Sun, X. Zhu, and Y. Fang, "Privacy and security for online social networks: challenges and opportunities", *Network, IEEE*, vol. 24, no. 4, pp. 13–18, 2010.
- [13] A. Sajid and H. Abbas, "Data privacy in cloud-assisted healthcare systems: State of the art and future challenges", *Journal of Medical Systems*, vol. 40, no. 6, pp. 1–16, 2016.
- [14] Zhengyuan Pang; Lifeng Sun; Zhi Wang; Erfang Tian; Shiqiang Yang, "A Survey of Cloudlet Based Mobile Computing", *International Conference on Cloud Computing and Big Data (CCBD)*, Pages: 268–275, Year: 2015.
- [15] Koustabh Dolui; Soumya Kanti Datta, "Comparision of Edge Computing Implementations: Fog Computing, Cloudlet and Mobile Edge Computing", *Global Internet of Things Summit (GloTS)*, Pages: 1–6, Year: 2017.
- [16] Aymen El Amraoui; Kaouthar Sethom, "Cloudlet Softwarization for Pervasive Healthcare", *30th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, Pages: 628–632, Year: 2016.
- [17] Praseetha V. M.; S. Vadivel, "Face Extraction using Skin Color and PCA Face Recognition in a Mobile Cloudlet Environment", *4th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud)*, Pages: 41–45, Year: 2016.

AUTHORS PROFILE

Dr. Anjali Chandavale

Dr. Anjali Chandavale is currently a Professor in Information Technology and handling the Responsibility of Centre of Excellence in Computing & IT applications at MIT College of Engineering, Pune, India. She has completed her doctoral study from University of Pune with College of Engineering, Pune, India as a research centre. She has wide experience of 19 years in the challenging and dynamic environment of education and 6 years in the industry. Her areas of interest are internet of things, internet security and computer vision.

Email address: c.anjali38@gmail.com

Anuja J. Gade

Pursuing M.E. (I.T.), B.E. (I.T.)

Anuja J. Gade currently pursuing M.E. from Savitribai Phule Pune University, Pune and obtained B.E. from Shivaji University, Kolhapur (Maharashtra), India in 2014. She is currently studying in Information Technology Department in MAEER's MIT College of Engineering, Pune, India. Her area of interest are cloud computing, internet of things.

Email address: anujagade08@gmail.com

Contact no: 9405542728

FiDooP: Parallel Mining of Frequent Itemsets Using MapReduce

Dr G Krishna Kishore¹
Computer Science and Engineering
V. R. Siddhartha Engineering College
Vijayawada, Andhra Pradesh, India
gkk@vrsiddhartha.ac.in

Suresh Babu Dasari²
Computer Science and Engineering
V. R. Siddhartha Engineering College
Vijayawada, Andhra Pradesh, India
dasarisuresh88@gmail.com

S. Ravi Kishan³
Computer Science & Engineering
V.R.Siddhartha Engineering College
Vijayawada, Andhra Pradesh
suraki@vrsiddhartha.ac.in

Abstract: Existing parallel digging calculations for visit itemsets do not have a component that empowers programmed parallelization, stack adjusting, information conveyance, and adaptation to non-critical failure on substantial bunches. As an answer for this issue, we outline a parallel incessant itemsets mining calculation called FiDooP utilizing the MapReduce programming model. To accomplish compacted capacity and abstain from building contingent example bases, FiDooP joins the incessant things Ultrametric tree, as opposed to ordinary FP trees. In FiDooP, three MapReduce occupations are actualized to finish the mining undertaking. In the essential third MapReduce work, the mappers autonomously disintegrate itemsets, the reducers perform mix activities by building little Ultrametric trees, and the genuine mining of these trees independently. We actualize FiDooP on our in-house Hadoop group. We demonstrate that FiDooP on the group is touchy to information dissemination and measurements, in light of the fact that itemsets with various lengths have diverse decay and development costs. To enhance FiDooP's execution, we build up a workload adjust metric to quantify stack adjust over the group's registering hubs. We create FiDooP-HD, an augmentation of FiDooP, to accelerate the digging execution for high-dimensional information investigation. Broad tests utilizing genuine heavenly phantom information exhibit that our proposed arrangement is productive and versatile.

Keywords - MapReduce, Frequent Itemsets Mining, Hadoop, Ultrametric, Celestial Spectral Data.

1. Introduction:

Visit Itemsets Mining (FIM) is a center issue in affiliation run mining (ARM), succession mining, and so forth. Accelerating the procedure of FIM is basic and basic, on the grounds that FIM utilization represents a critical segment of mining time because of its high calculation and information/yield (I/O) power. At the point when datasets in present day information mining applications turn out to be too much substantial, successive FIM calculations running on a singlemachine experience the ill effects of execution disintegration. To address this issue, we explore how to perform FIM utilizing MapReduce a broadly embraced programming model for handling huge datasets by misusing the parallelism among registering hubs of a group. We demonstrate to disseminate an extensive dataset over the group to adjust stack over all bunch hubs, in this manner enhancing the execution of parallel FIM.

2. LITERATURE REVIEW

Data mining faces a lot of challenges in the big data era. Association rule mining algorithm is not sufficient to process large data sets. Apriori algorithm has limitations like the high I/O load and low performance. The FP-Growth algorithm also has certain limitations like less internal memory. Mining the frequent itemset in the dynamic scenarios is a challenging task. A parallelized approach using the MapReduce framework is also used to process large data sets. The most efficient the recent method is the FiDooP using Ultrametric tree (FIUT) and MapReduce programming model. FIUT scans the database only twice. FIUT has four advantages. First: I reduces the I/O overhead as it scans the database only twice. Second: only frequent itemsets in each transaction are inserted as nodes for compressed storage. Third: FIU is improved way to partition database, which

significantly reduces the search space. Fourth: frequent itemsets are generated by checking only leaves of tree rather than traversing entire tree, which reduces the computing time. The mining of frequent itemsets is a basic and essential work in many data mining applications. Frequent itemsets extraction with frequent pattern and rules boosts the applications like Association rule mining, correlations also in product sale and marketing. In extraction process of frequent itemsets there are number of algorithms used like FP-growth, E-clat etc. But unfortunately these algorithms are inefficient in distributing and balancing the load, when it comes across massive data. Automatic parallelization is also not possible with these algorithms. To defeat these issues of existing algorithms there is need to construct an algorithm which will support the missing features, such as automatically parallelization, balancing and good distribution of data. This paper is focusing on an efficient methodology to extract frequent itemsets with the popular MapReduce approach. This new methodology consist an algorithm which is build using Modified Apriori algorithm, called as Frequent Itemset Mining using Modified Apriori (FIMMA) Technique. This methodology works with three mappers, independently and concurrently by using the decompose strategy. The result of these mappers will be given to the reducers using the hash table method. Reducer gives the top most frequent itemsets.

3. Proposed System

In Proposed System a new data partitioning method to well balance computing load among the cluster nodes; we develop FiDooop-HD, an extension of FiDooop, to meet the needs of high dimensional data processing.

Step 1: Count the occurrence of each item.

| Item | Occurrence / Frequency |
|------|------------------------|
| 1 | 3 |
| 2 | 3 |
| 3 | 2 |
| 4 | 5 |
| 5 | 4 |
| 6 | 3 |
| 7 | 1 |
| 8 | 1 |
| 9 | 2 |
| 0 | 2 |

Figure 3.1:Frequency of each item

Step 2: We start making pairs out of the frequent itemsets we got in the above step.

| ItemPairs |
|-----------|
| 12 |
| 14 |
| 15 |
| 16 |
| 24 |
| 25 |
| 26 |
| 45 |
| 46 |
| 56 |

Figure 3.2:Frequent item sets pairs.

Step 3: After getting the frequent Item Pairs, we start counting the occurrence of these pairs in the Transaction Set.

| ItemPairs | Occurrence / Frequency |
|-----------|------------------------|
| 12 | 1 |
| 14 | 2 |
| 15 | 2 |
| 16 | 1 |
| 24 | 3 |
| 25 | 3 |
| 26 | 2 |
| 45 | 4 |
| 46 | 3 |
| 56 | 2 |

Figure 3.3:Frequency of itemset pairs

Step 4: Make combinations of triples using the frequent Item pairs.

To make triples, the rule is: IF 12 and 13 are frequent, then the triple would be **123**. Similarly, if 24 and 26 then triple would be **246**.

So, using the above logic and our Frequent Item Pairs table, we get the below triples:

| ItemTriples |
|-------------|
| 245 |
| 456 |

Figure 3.4:Frequent itemset triplets.

Step 5: Get the count of the above triples (Candidates).

| ItemTriples | Occurrence / Frequency |
|-------------|------------------------|
| 245 | 3 |
| 456 | 2 |

Figure 3.5:Frequency of itemsets triplets.

After, this, if we can find quartets, then we find those and count their occurrence/frequency.

If we had 123, 124, 134, 135, 234 and we wanted to generate a quartet then it would be 1234 and 1345. And after finding quartet we would have again got their count of occurrence /frequency and repeated the same also, until the Frequent ItemSet is null.

Thus, the frequent ItemSets are:

- Frequent Itemsets of Size 1: 1, 2, 4, 5, 6
- Frequent Itemsets of Size 2: 14, 24, 25, 45, 46
- Frequent Itemsets of Size 3: 245

3.1 METHODOLOGY

In Proposed System a new data partitioning method to well balance computing load among the cluster nodes; we develop FiDooop-HD, an extension of FiDooop, to meet the needs of high dimensional data processing. FiDooop is efficient and scalable on Hadoop clusters.

The proposed system involves the following steps:

- Load the data base into the system.
- Perform mining on all datasets of the database.
- Calculate the support values and confidence values of the datasets.
- Sort the elements based on their support values.
- Set the threshold support value.
- Extract the elements with support values above threshold.

Approach

- 1) Finding the Frequent Items: During the first step, the vertical database is divided into equally sized blocks (shards) and distributed to available mappers. Each mapper extracts the frequent singletons from its shard. In the reduce phase, all frequent items are gathered without further processing.
- 2) k-FIs Generation: In this second step, Pk, the set of frequent itemsets of size k, is generated. First, frequent singletons are distributed across m mappers. Each of the mappers finds the frequent k-sized supersets of the items by running Eclat to level k. Finally, a reducer assigns Pk to a new batch of m mappers. Distribution is done using Round-Robin.
- 3) Subtree Mining: The last step consists of mining the prefix tree starting at a prefix from the assigned batch using Eclat. Each mapper can complete this step independently since sub-trees do not require mutual information.

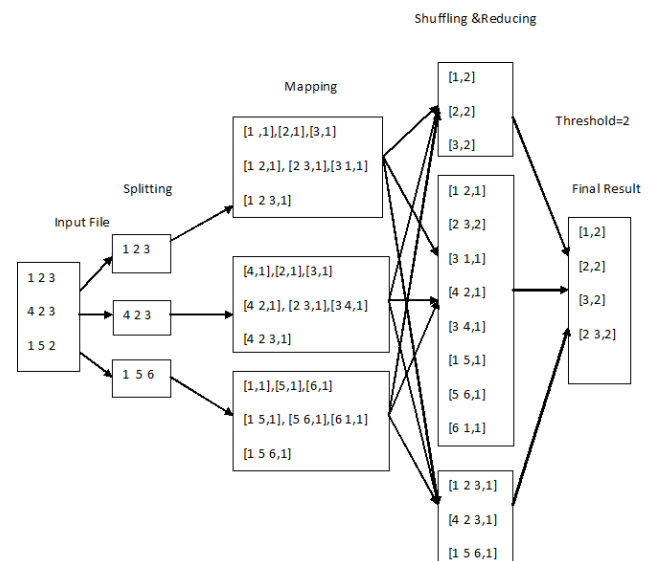


Figure 3.1.1 Map Reduce process

4. IMPLEMENTATION:

Data set: Groceries data set in csv format.

INPUT: Transactions dataset i.e groceries dataset.

OUTPUT: Frequent itemsets

There are three modules in the proposed system. They are as follows:

MODULE 1:

The first mapper program would mine the transaction database by removing infrequent sets. This output from the map is given to reducer as an input which would order the frequent itemsets in descending order and would build a FP tree.

Algorithm:

Input: minsupport, DBi;

Output: FP tree

1. function MAP(key offset, values DBi)
2. //T is the transaction in DBi
3. for all T do
4. items \leftarrow split each T;
5. for all item in items do 1. count++ 2. end for
6. output(item, count);
7. end for
8. end function
10. reduce input: (itemset, count)
11. function REDUCE(key item, values count)
12. Items=sort(itemset, count) /*sorts the items in descending order*/
13. fptree_generation(items); /*generates FP tree */
14. end function

MODULE 2:

The second map - reducer program takes the output from the second reducer , which would recursively processes the data and generates a minimum 2 Item sets using the FiDooPHD algorithm.

Algorithm:

Input: List,

Output:-FP Tree

1. function MAP(List)
2. // M is the size of the List 2. for all (k is from M to 2) do
3. for all (k-itemset in List) do
4. decompose(k-itemset, k-1, (k-1)-itemsets); /*Each k-itemset is only decomposed into (k-1)-itemsets */
5. (k-1)-file \leftarrow the decomposed (k-1)-itemsets
6. union the original (k-1)-itemsets in (k-1)-file; 2. for all (t-itemset in (k-1)-file) do 3. t -FP-tree \leftarrow t-FP-tree generation(local-FPtree,t itemset);
8. output(t, t-FP-tree);
9. end for
10. end for
11. end for

12. end function

5. OUTPUT:

The following diagrams shows the implementation of Fidoop and display of frequent itemsets for the given datasets.

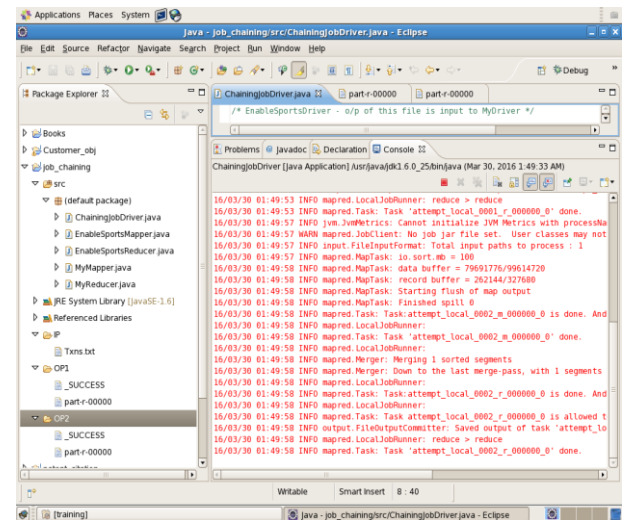


Figure 5.1 Execution of Fidoop

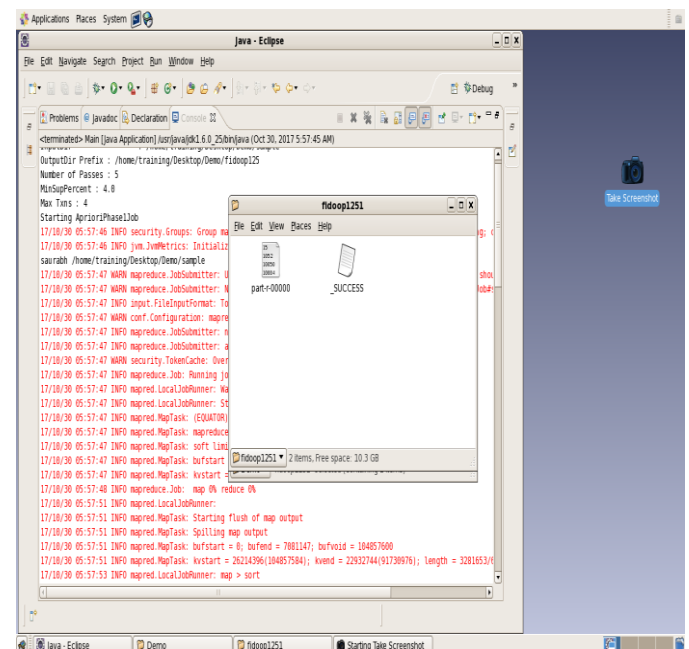


Figure 5.2: Generation of Output File and Success File

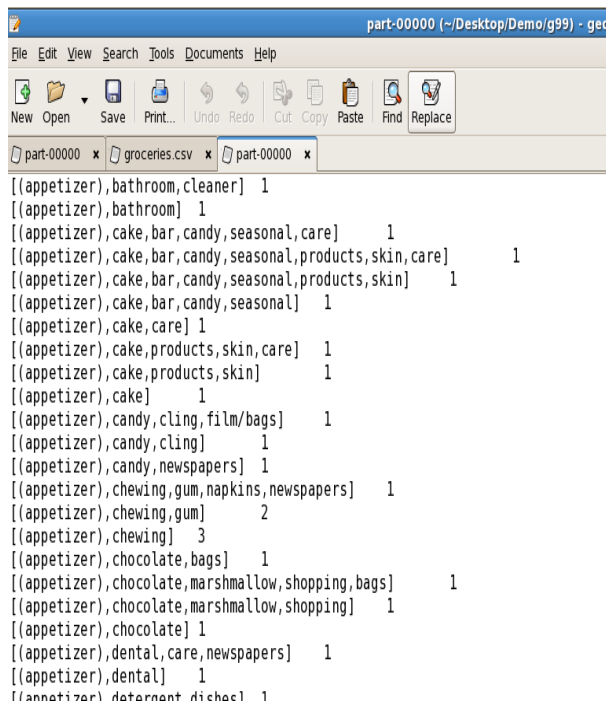


Figure 5.3: Display of Frequent Item Sets

6. CONCLUSION AND FUTURE WORK

To mitigate high communication and reduce computing cost in MapReduce-based FIM algorithms, we developed FiDooP-DP, which exploits correlation among transactions to partition a large dataset across data nodes in a Hadoop cluster. FiDooP-DP is able to partition transactions with high similarity together and group highly correlated frequent items into a list.

7. REFERENCES

- 1) Shreedevi C Patil "A Survey on Parallel Mining of frequent Itemsets in MapReduce", International Journal of Innovative Research in Computer and Communication Engineering, Volume 4, Issue-6, June, 2016.
- 2) Prajakta G. Kulkarni, S.R.Khonde "An Improved Technique Of Extracting Frequent Itemsets From Massive Data Using MapReduce", International Journal of Engineering and Technology, Volume-9, July, 2017.
- 3) ShivaniDeshpande, HarshitaPawar, Amruta Chandras, AmolLanghe "Data Partitioning in Frequent Itemset Mining on Hadoop Clusters", International Research Journal of Engineering and Technology (IRJET), Volume: 03 Issue: 11, November, 2016.
- 4) Divya.M.G, Nandini.K, Priyanka.K.T, V and ana.B "Weighted Itemset Mining from Big Data using Hadoop", International Journal of Advanced Networking & Applications, ISSN: 0975-0282, February, 2016.
- 5) Roger Pressman, titled "Software Engineering - a practitioner's approach", Fifth Edition.
- 6) Herbert Schildt, titled "The Complete Reference Java", Seventh Edition.
- 7) Tom White, titled "Hadoop: The Definitive Guide", Third Edition.
- 8) Robin Nixon, titled "Learning PHP, MySQL & JavaScript".
- 9) J. des Riviere, J. Wiegand "Eclipse: A platform for integrating development tools", IBM SYSTEMS JOURNAL, Volume: 43, NO 2, 2004.

Pleomorphism in Cervical Nucleus: A Review

Jing Rui Tang, *Member, IEEE*

Abstract—Cervical cancer is the fourth most common cancer among women worldwide but the disease is preventable. Papanicolaou test enables detection of the precancerous cells on the cervix based on the examination of slide under the microscope. Cervical cancer is graded based on the morphological changes on the cells and pleomorphism is one of the prominent characteristic. This paper briefly reviews recent publications that work directly or indirectly on pleomorphism. Based on the review, it is noticed that some features for nuclear shape were widely used, including area, perimeter, major and minor axis lengths, circularity and eccentricity. As a prominent feature which could be identified easily during examination of slides, future works could take into consideration on how the human experts define pleomorphism. The correlation between those computed features and how human eyes recognize shape variation could be studied. Quantification of pleomorphism is necessary to reduce vagueness and ambiguity in justifying pleomorphism.

Index Terms—Cervical cancer, feature extraction, nucleus, pleomorphism.

I. INTRODUCTION

CERVICAL cancer is the fourth most common cancer among women worldwide, with an estimated of more than 250,000 deaths yearly. There were 528,000 new cases worldwide in 2012 and approximately 84% occurred in less developed countries [1]. Cervical cancer is in fact preventable and highly treatable if detected early [2-4]. Screening of cervical cancer, more commonly known as Papanicolaou test (i.e., Pap test), identify the precancerous or cancerous cells on the cervix based on the examination of slide under the microscope and thus prevents further progression of the cells into a more invasive stage. Cervical cancer is graded based on the morphological changes on the cells [5-7]. In a review by [8], the authors studied both the concepts and terminology employed for cervical precancerous morphological changes and its relationship with the natural history through information from cervical screening for better understanding of the complex link between cytological and histological diagnosis and the natural history of cervical precancerous stage. By correlating the cervical cytology report with the histopathological diagnosis, a comparative study using 3438 Pap smear from the health centres in Theni district, India analyzed the accuracy of the cervical cytology report based on the Bethesda system [9]. Some of the visible characteristics of cervical cells as they

progress from normal to abnormal stage include changes in color (i.e., the nucleus become darker in color due to the presence of highly stained chromatin), changes in shape (i.e., pleomorphism, whereby the nuclear shape becomes bizarre as the nucleus can hardly retained its shape due to uncontrollable division) and changes in nuclear size (i.e., the nucleus becomes larger) [4, 10-13]. A Pap test result is reported according to either Bethesda System for Reporting Cervical Cytology [14] or to the British Society for Clinical Cytologists (BSCC) Terminology [15]. With the advancement in technology, many cervical cancer screening systems have been developed for the automation of the screening process. Multiple features are used for classification with several different types of classifiers such as support vector machine and artificial neural network [16-19]. One of the criteria in both the reporting standards is the changes in the shape of the cell nuclei. In this study, we are focusing on the shape of the nucleus since this feature appeared to be one of the most significant visible characteristics. The study will first review previous work on cervical cell shape analysis, follow by the challenges and suggestions for future work.

II. MEASURING PLEOMORPHISM

Various studies have reported ways of analyzing shape [20-22]. Shape as a diagnostic characteristic is not something new in medical field. As early as in 1978, shape-oriented parameters were computed but quantification of shape only performed for cytoplasm [23]. Also, a robust deformable segmentation framework which integrated sparse shape composition was proposed in [24]. The performance of the proposed approach was validated via lung localization in X-ray, three-dimensional images of liver in positron emission tomography-computed tomography and rat cerebellum segmentation in magnetic resonance microscopy. Significance of the nuclear shape as one of the observable morphological change in cervical nuclei as the cells progress from normal to abnormal stage is demonstrated elsewhere [6, 14, 25]. However, computational of shape feature might be expensive and time consuming [26-28].

In the proposed Median M-Type Radial Basis Function neural network [29], nine features were extracted. Features related to the measurement of nuclear shape included nuclear perimeter and circularity. Here, nuclear perimeter is defined as the summation of the pixels which form the outline of the nucleus. In a study of approximately forty methods for shape feature extraction, the authors pointed out that shape could be described from different aspects [30]. Apart from some widely

Jing Rui Tang is with the Faculty of Technical and Vocational, Sultan Idris Education University, Malaysia. (e-mail: tangjingrui@ftv.upsi.edu.my).

used parameters such as eccentricity and circularity, other shape parameters included the center of gravity, axis of least inertia, bending energy, elliptic variance, rectangularity and convexity.

With the observation that all the nuclear sizes lie within a certain range with the compact and normally smooth nuclear shape, Hough transform was implemented to localized the cervical cell nuclei [31]. Further processing was performed using a level set algorithm and the algorithm was tested with 207 images. In [32], the segmented nuclei were expected to be elliptical in shape and six shape features were computed, including minor axis length, major axis length, eccentricity, equivalent diameter, perimeter and circularity. The proposed automated approach detected candidate nuclei using morphological image reconstruction while nuclei boundary was segmented by watershed transform.

For the segmentation of nucleus and cytoplasm in cervical smear images, Radiating Gradient Vector Flow Snake was proposed by [33]. The proposed approach which introduced a novel edge map computation method together with refinement based on stack demonstrated great potential to locate the obscure boundaries including those interferences near the regions of nucleus and cytoplasm. By incorporating both local and global schemes in the proposed graph cut approach for the segmentation of nuclei and cytoplasm, simulation results using twenty one cervical cell images by Zhang et al. returned accuracy of F-measure of 88.4% for abnormal nuclei binarization. They used morphological and gradient features to separate the touching nuclei that fulfill the criteria as touching-nuclei clump (i.e., via computation of roundness and shape factor) [12].

In a proposed method for automatic cervical cancer cell segmentation and classification, a single-cell image is divided into three regions (i.e., the nucleus, cytoplasm, and background), using the fuzzy C-means clustering technique and the results were compared with hard C-means clustering and watershed technique [4]. Using the nine features extracted (i.e., six features extracted from the nucleus and the remaining three features from cytoplasm), five of the six nucleus-related features are highly correlated with the shape. The features used are as followed: area of nucleus (1), compactness of nucleus (2), major axis of nucleus, M_{major} (i.e., the length of the major axis of the ellipse which totally encloses the region of the nucleus), minor axis of nucleus, M_{minor} (i.e., the length of the minor axis of the ellipse which totally encloses the region of the nucleus) and aspect ratio of nucleus (3).

$$Area, A = \sum_{i=1}^n i \quad (1)$$

where n is the total number of pixels in the nucleus region.

$$Compactness, Com = \frac{(Perimeter, P)^2}{A} \quad (2)$$

where P is total number of pixels that forms the boundary of the nucleus region.

$$AspectRatio, AR = \frac{W}{H} \quad (3)$$

where W and H are the width and height of the nucleus region,

respectively.

In the review article [11], the shape is defined by several measurements, including the length of the major and minor axes, symmetry and circularity. The importance of the feature selection in resulting good classification results is discussed in [34]. The study proposed a nominated texture based cervical cancer classification system whereby seven feature sets that contained of twenty four features were used for classification, including relative size of nuclei and cytoplasm, gray level co-occurrence matrix features, Tamura features and edge orientation histogram. Here, one of the Tamura's texture features, the coarseness, could be seen as the nuclei shape information.

In their attempt to quantify features and further detect abnormal cervical squamous epithelial cells, Mingzhu Zhao et al. extracted descriptors based on morphology, color and texture features of cervical squamous epithelial cells [35]. They presented the morphological difference degree in two parts, namely size and shape difference degrees. The shape difference degree, mainly to describe the heteromorphic features of nucleus, was depicted in two pathology-related ways. The first way takes into account the circularity and the compactness of the nucleus while the second way deals with the descriptor of nuclear boundary.

Using different perspective for analysis as compared with other approaches, two techniques were proposed for the evaluation of nuclear membrane irregularity [25]. The first technique imposed different penalty weighting so that a more irregular nuclear membrane will receive higher penalty while the second technique computed how much the nuclear membrane contour deviated from the mean and also median values of nuclear membrane contour.

By combining shape detection and artificial neural network, a proposed cervical nuclei extraction method could manage multiscale information and returned accurate results [36]. In order to discriminate nucleus and non-nucleus, three different features (i.e., intensity, shape and texture) were used. A total of seven shape features were computed, including area, perimeter, circularity (4), equivalent diameter (5), major axis length, minor axis length, eccentricity (6) and number of curvature sign changes.

$$Circularity, Cir = \frac{4\pi A}{P^2} \quad (4)$$

$$EquivalentDiameter, ED = \frac{4A}{\pi} \quad (5)$$

$$Eccentricity, Ecc = \frac{\sqrt{M_{major} - M_{minor}}}{M_{major}} \quad (6)$$

Using the proposed two-level cascade classifier, twenty eight dimensional features in morphology and texture were used for the classification of the cervical cancer cells and achieved 1.44% for both the false positive and false negative rates [37]. The morphologic features used to describe the shape of the nucleus, included the area, circularity, distance (7), sigma (8),

roundness (9), sides (10) and many others [38].

$$Distance, Dist = \frac{\sum_{u,v} (\|p_o - p_{(u,v)}\|)}{A} \quad (7)$$

$$Sigma, \sigma = \sqrt{\frac{\sum_{u,v} (\|p_o - p_{(u,v)}\| - Dist)^2}{A}} \quad (8)$$

where p_o and $p_{(u,v)}$ are the mean and pixel values in the position (u,v) in the area of interest.

$$Roundness, Rou = 1 - \frac{\sigma}{Dist} \quad (9)$$

$$Sides, Sid = 1.4111 * \left(\frac{Dist}{\sigma}\right)^{0.4724} \quad (10)$$

Via the proposed Markov random field segmentation framework, [39] treated the input cervical cell images as an undirected probabilistic graphical model. A total of thirteen features including the shape of superpixel patches were used for separation of nuclei, cytoplasm and background.

The term 'pleomorphism' was specifically mentioned in [40]. A technique was proposed for the objective measurement of pleomorphism based on the widely used gray level co-occurrence matrix (GLCM). The proposed technique, named as Cell Feature Level Co-occurrence Matrix, extracted sixteen nuclear shape related features. Also, [41] studied the nuclear pleomorphism in various stages of oral carcinogenesis. It is stated that nuclear pleomorphism that presents in round, oval, spindle, elongated fiber and irregular shapes were often seen during different stages of carcinogenesis.

In the study of high-grade squamous intraepithelial lesions (HSILs) demonstrating bizarre cytological appearances, or named as 'pleomorphic HSIL', out of the nineteen cases, 16% of them have superficially invasive squamous cell carcinoma. Nonetheless, their findings reveal that pleomorphism in HSIL sometimes do not necessarily represent more aggressive biological behavior. It could indicate a degenerative phenomenon and hence the authors suggested that more aggressive clinical management is not necessarily for pleomorphic HSIL compared to the conventional HSIL but required bigger-scaled long-term investigations [42].

In [10], the proposed graph-search based method successfully took into account the nuclei shape information during the graph construction, resulting in a superior-performance segmentation method for abnormal cervical nuclei. Bora K. et al. proposed an intelligent cervical-dysplasia-detection system that classified the cervical dysplasia into bi-class (i.e., normal and abnormal) and tri-class (i.e., NILM, LSIL and HSIL) using shape, texture and color features. The shape descriptors for the nucleus are area, perimeter, eccentricity, compactness and circularity [43].

In [44], Tareef A. et al. proposed a two-stage segmentation approach which incorporated shape and appearance features in superpixel representation level. During the first stage, support vector machine to classify regions of the image into nuclei, cellular clusters, and background based on the superpixel-based features of local discriminative shape and appearance cues. The

second stage demonstrated the proposed shape deformation framework which forms the cytoplasmic shape of every overlapping cell followed by shape refinement using Distance Regularized Level Set Evolution model. Simulation results revealed that the proposed approach was capable to separate touching and heavily-overlapping cells from large clusters.

In a recent work to segment abnormal cervical cell nuclei [10], graph-search based segmentation was integrated with a two-dimensional dynamic programming approach to improve cell nucleus segmentation. Nuclear shape, border and regional information together with nuclear context prior constraints were employed and the results were validated by Herlev dataset and H&E stained manual liquid-based cytology dataset with comparison with five state-of-the-art techniques.

Although many works on classification as well as segmentation of cervical cell images have been published, it is noticed that most of the works employed the same features for nuclei shape. Some widely used features for nuclear shape include area, perimeter, major axis length, minor axis length, circularity and eccentricity. As a prominent feature which could be identified easily during the examination of slides, intriguingly limited studies have focused on nuclei shape.

III. CHALLENGES AND FUTURE WORKS

Justification of shape features could be highly subjective. Based on findings in Section II, it is found that very limited work has taken into consideration the human experts' perception in perceiving nuclear shape. Also, the context of 'pleomorphism' could vary depending on the background of individual pathologist and cytotechnologist, particularly on the degree of pleomorphism. Hence, future work should take into account on how the human experts define pleomorphism. Further, correlation between those computed parameters such as area and perimeter with how human eyes recognize shape variation could be studied.

Furthermore, quantification of pleomorphism could be placed into focus for future work, whereby the term could be transformed into a measurable parameter. Standardization of the term not only helps in reducing the vagueness and ambiguity, it also contributes to reducing the miscommunication as well as misconception and hence indirectly promotes more accurate and consistent Pap test results.

IV. CONCLUSION

Cervical cancer is graded based on the morphological changes on the cells. Pleomorphism is known as one of the observable morphological changes that are prominent. This review paper studied recent publications which worked directly or indirectly on pleomorphism. Some nuclear-shape-related-features such as area, perimeter and eccentricity were widely used. Future works could study the correlation between the computed

features with the human's perception regarding shape variation. Quantification of pleomorphism is important to minimize the vagueness due to subjective justification of pleomorphism.

REFERENCES

- [1] W. H. O. Report, "Latest world cancer statistics - Global cancer burden rises to 14.1 million new cases in 2012: Marked increase in breast cancers must be addressed," World Health Organization, Lyon/Geneva 2013.
- [2] World Health Organization, "Human papillomavirus (HPV) and cervical cancer," WHO Media Centre 2016.
- [3] J. R. Tang, N. A. M. Isa, and E. S. Ch'ng, "Segmentation of cervical cell nucleus using Intersecting Cortical Model optimized by Particle Swarm Optimization," in *2015 IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, 2015, pp. 111-116.
- [4] T. Chankong, N. Theera-Umporn, and S. Auephanwiriyakul, "Automatic cervical cell segmentation and classification in Pap smears," *Computer Methods and Programs in Biomedicine*, vol. 113, pp. 539-556, 2014.
- [5] A. H. Fischer, C. Zhao, Q. K. Li, K. S. Gustafson, I. E. Eltoum, R. Tambouret, *et al.*, "The cytologic criteria of malignancy," *Journal of Cellular Biochemistry*, vol. 110, pp. 795-811, 2010.
- [6] P. Dey, "Cancer nucleus: Morphology and beyond," *Diagnostic Cytopathology*, vol. 38, pp. 382-390, 2010.
- [7] L. G. Koss and M. R. Melamed, *Koss' diagnostic cytology and its histopathologic bases*. New York: Lippincott Williams & Wilkins, 2006.
- [8] D. Jenkins, "Histopathology and cytopathology of cervical cancer," *Disease Markers*, vol. 23, pp. 199-212, 2007.
- [9] D. Suryamoorthy, K. Duraisamy, and R. Ramakrishnan, "A comparative study of cervix cytology smears with histopathological findings," *Journal of Evolution of Medical And Dental Sciences* vol. 6, pp. 927-930, 2017.
- [10] L. Zhang, H. Kong, S. Liu, T. Wang, S. Chen, and M. Sonka, "Graph-based segmentation of abnormal nuclei in cervical cytology," *Computerized Medical Imaging and Graphics*, vol. 56, pp. 38-48, 2017.
- [11] Y. Jusman, S. C. Ng, and N. A. Abu Osman, "Intelligent screening systems for cervical cancer," *The Scientific World Journal*, vol. 2014, p. 15, 2014.
- [12] L. Zhang, H. Kong, C. T. Chin, S. Liu, Z. Chen, T. Wang, *et al.*, "Segmentation of cytoplasm and nuclei of abnormal cells in cervical cytology using global and local graph cuts," *Computerized Medical Imaging and Graphics*, vol. 38, pp. 369-380, 2014.
- [13] B. Sokouti, S. Haghipour, and A. D. Tabrizi, "A pilot study on image analysis techniques for extracting early uterine cervix cancer cell features," *Journal of Medical Systems*, vol. 36, pp. 1901-1907, June 01 2012.
- [14] R. Nayar and D. C. Wilbur, *The Bethesda System for reporting cervical cytology*: Springer, 2015.
- [15] K. J. Denton, A. Herbert, L. S. Turnbull, C. Waddell, M. S. Desai, D. N. Rana, *et al.*, "The revised BSCC terminology for abnormal cervical cytology," *Cytopathology*, vol. 19, pp. 137-157, 2008.
- [16] A. Bhargava, P. Gairola, G. Vyas, and A. Bhan, "Computer aided diagnosis of cervical cancer using HOG features and multi classifiers," in *Intelligent Communication, Control and Devices. Advances in Intelligent Systems and Computing*, Singapore, 2018, pp. 1491-1502.
- [17] H. A. Almubarak, R. J. Stanley, R. Long, S. Antani, G. Thoma, R. Zuna, *et al.*, "Convolutional neural network based localized classification of uterine cervical cancer digital histology images," *Procedia Computer Science*, vol. 114, pp. 281-287, 2017.
- [18] E. T. Tan, J. R. Tang, and N. A. M. Isa, "Applying design of experiment to optimise artificial neural network for classification of cervical cancer," *Journal of Engineering Science*, vol. 12, pp. 65-75, 2016.
- [19] M. A. Devi, S. Ravi, J. Vaishnavi, and S. Punitha, "Classification of cervical cancer using artificial neural networks," *Procedia Computer Science*, vol. 89, pp. 465-472, 2016.
- [20] L. Pishchulin, S. Wuhrer, T. Helten, C. Theobalt, and B. Schiele, "Building statistical shape spaces for 3D human modeling," *Pattern Recognition*, vol. 67, pp. 276-286, 2017/07/01/ 2017.
- [21] E. Nelson, J. Hall, P. Randolph-Quinney, and A. Sinclair, "Beyond size: The potential of a geometric morphometric analysis of shape and form for the assessment of sex in hand stencils in rock art," *Journal of Archaeological Science*, vol. 78, pp. 202-213, 2017.
- [22] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models - Their training and application," *Computer Vision and Image Understanding*, vol. 61, pp. 38-59, 1995.
- [23] J. Holmquist, E. Bengtsson, O. Eriksson, B. Nordin, and B. Stenkvist, "Computer analysis of cervical cells. Automatic feature extraction and classification," *Journal of Histochemistry & Cytochemistry*, vol. 26, pp. 1000-1017, 1978.
- [24] S. Zhang, Y. Zhan, and D. N. Metaxas, "Deformable segmentation via sparse representation and dictionary learning," *Medical Image Analysis*, vol. 16, pp. 1385-1396, 2012.
- [25] J. R. Tang, N. A. Mat Isa, and E. S. Ch'ng, "Evaluating nuclear membrane irregularity for the classification of cervical squamous epithelial cells," *PLoS ONE*, vol. 11, p. e0164389, 2016.
- [26] S. Ali, R. Veltri, J. I. Epstein, C. Christudass, and A. Madabhushi, "Selective invocation of shape priors for deformable segmentation and morphologic classification of prostate cancer tissue microarrays," *Computerized Medical Imaging and Graphics*, vol. 41, pp. 3-13, 2015.
- [27] S. Ali and A. Madabhushi, "An integrated region-, boundary-, shape-based active Contour for multiple object overlap resolution in histological imagery," *IEEE Transactions on Medical Imaging*, vol. 31, pp. 1448-1460, 2012.
- [28] M. Rousson and N. Paragios, "Shape priors for level set representations," *Berlin, Heidelberg*, 2002, pp. 78-92.
- [29] M. E. Gómez-Mayorga, F. J. Gallegos-Funes, J. M. De-la-Rosa-Vázquez, R. Cruz-Santiago, and V. Ponomaryov, "Diagnosis of cervical cancer using the Median M-Type Radial Basis Function (MMRBF) neural network," *Berlin, Heidelberg*, 2009, pp. 258-267.
- [30] M. Yang, K. Kpalma, and J. Ronsin, "A survey of shape feature extraction techniques," in *Pattern Recognition*, Y. Peng-Yeng, Ed., ed: IN-TECH, 2008, pp. 43-90.
- [31] C. Bergmeir, M. García Silvente, and J. M. Ben fez, "Segmentation of cervical cell nuclei in high-resolution microscopic images: A new algorithm and a web-based software framework," *Computer Methods and Programs in Biomedicine*, vol. 107, pp. 497-512, 2012.
- [32] M. E. Plissiti, C. Nikou, and A. Charchanti, "Combining shape, texture and intensity features for cell nuclei extraction in Pap smear images," *Pattern Recognition Letters*, vol. 32, pp. 838-853, 2011.
- [33] K. Li, Z. Lu, W. Liu, and J. Yin, "Cytoplasm and nucleus segmentation in cervical smear images using Radiating GVF Snake," *Pattern Recognition*, vol. 45, pp. 1255-1264, 2012.
- [34] E. J. Mariarputham and A. Stephen, "Nominated texture based cervical cancer classification," *Computational and Mathematical Methods in Medicine*, vol. 2015, p. 10, 2015.
- [35] M. Zhao, L. Chen, L. Bian, J. Zhang, C. Yao, and J. Zhang, "Feature quantification and abnormal detection on cervical squamous epithelial cells," *Computational and Mathematical Methods in Medicine*, vol. 2015, p. 9, 2015.
- [36] D. Garcia-Gonzalez, M. Garcia-Silvente, and E. Aguirre, "A multiscale algorithm for nuclei extraction in pap smear images," *Expert Systems with Applications*, vol. 64, pp. 512-522, 2016.
- [37] J. Su, X. Xu, Y. He, and J. Song, "Automatic detection of cervical cancer cells by a two-level cascade classification system," *Analytical Cellular Pathology*, vol. 2016, p. 11, 2016.
- [38] L. Zhang, S. Chen, T. Wang, Y. Chen, S. Liu, and M. Li, "A practical segmentation method for automated screening of cervical cytology," in *2011 International Conference on Intelligent Computation and Bio-Medical Instrumentation*, 2011, pp. 140-143.
- [39] L. Zhao, K. Li, M. Wang, J. Yin, E. Zhu, C. Wu, *et al.*, "Automatic cytoplasm and nuclei segmentation for color cervical smear image using an efficient gap-search MRF," *Computers in Biology and Medicine*, vol. 71, pp. 46-56, 2016.

- [40] A. Saito, Y. Numata, T. Hamada, T. Horisawa, E. Cosatto, H.-P. Graf, *et al.*, "A novel method for morphological pleomorphism and heterogeneity quantitative measurement: Named cell feature level co-occurrence matrix," *Journal of Pathology Informatics*, vol. 7, pp. 36-36, January 1, 2016 2016.
- [41] A. Mohanta and P. K. Mohanty, "Nuclear pleomorphism-based cytopathological grading in human oral neoplasm," *Russian Open Medical Journal*, vol. 6, 2017.
- [42] C. J. R. Stewart, "High-grade squamous intraepithelial lesion (HSIL) of the cervix with bizarre cytological appearances ('pleomorphic HSIL'): a review of 19 cases," *Pathology*, vol. 49, pp. 465-470, 2017.
- [43] K. Bora, M. Chowdhury, L. B. Mahanta, M. K. Kundu, and A. K. Das, "Automated classification of Pap smear images to detect cervical dysplasia," *Computer Methods and Programs in Biomedicine*, vol. 138, pp. 31-47, 2017/01/01/ 2017.
- [44] A. Tareef, Y. Song, W. Cai, H. Huang, H. Chang, Y. Wang, *et al.*, "Automatic segmentation of overlapping cervical smear cells based on local distinctive features and guided shape deformation," *Neurocomputing*, vol. 221, pp. 94-107, 2017.

Comparative Study of diverse API Perspective of Spatial Data

Zainab Tariq¹ and Umesh Chandra²

¹Research Associate, Computer Science, Glocal University, Saharanpur

²Assistant Professor, Computer Science, Banda University of Agriculture and Technology, Banda

Abstract

Application Programming interface (API) is cipher which enables two software programs to interact with each other. There are number of APIs available in the market like Map APIs, Facebook API, Amazon API, YouTube API, Flickr API, Twitter API, Word press API, Drop box API etc. Present study deals with APIs related to Maps that is Map API. Map APIs are used for navigation purposes, for finding real time locations like railways, road transport etc., and static locations like eateries, movie theatres, shopping malls, book stores and what not. This study emphasises the differences among various Map APIs on the basis of architecture used, technology followed, platform, programming language, open source and android support. This review will help one to choose a particular Map API and take the advantage by using its specific functionalities.

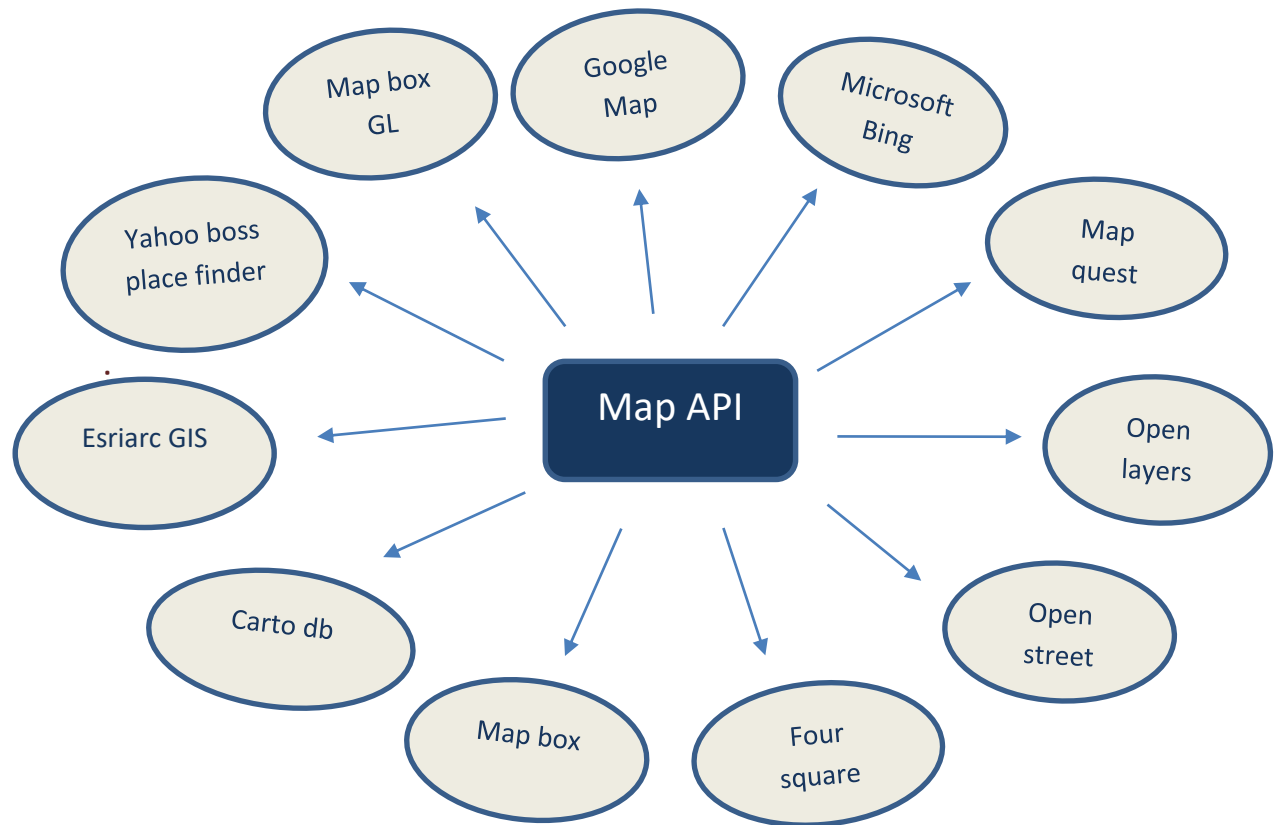
Keywords : Map API, navigation, architecture, technology.

1.1 Introduction

API is a structure that helps us to build an interface for application programs to be run on various platforms like android, IOS and other platforms. An API equips us with certain subprograms, tools and procedures for developing application software. Moreover, API equips us with some functions and classes which assist us to keep away from composing of low level cipher to carry out certain things. API is a set of functions called by some programming language and not to be confused with itself being programming language. A number of APIs exist today for various purposes – API for web Maps like Google Maps, API for social media like Facebook, API for E-commerce like Amazon, API for online videos like YouTube, API for content management like WordPress, API for photo sharing like Flickr, API for file or document management like Drop box, and so many others.

Web Mapping API is the API designed particularly for creating maps on web. These APIs comprise of classes for layers and maps which saves one from composing the low level cipher for exhibiting interactive images of map and representing it with another layer. There is a number of Web mapping APIs created like Google Map API, Microsoft Bing API, Yahoo boss place finder API, Map box, Open layers, Open street, etc.

Depending upon the characteristics, the API is chosen and worked upon. Discussed are the various differences based on parameters like architecture used, technology used, programming language and other characteristics of this Web mapping APIs in this study.



1.2 Architecture used in various APIs

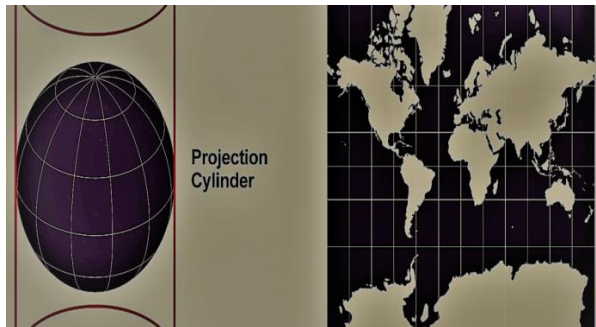
Architecture of API implies to build a base for developing an application which is robust and scalable. The requirement here is that the architecture should be able to enhance the interactivity of the software or service. Architecture of API ensures data to be frisky, application to be aware of the conditions and prognostic and should be able to administer locally.

Basically the idea of architecture is to integrate loosely coupled services in order to develop an interface which is easily accessible through web, android and other platforms. Enterprise Service Bus (ESB); JavaScript Object Notation (JSON); Representational State Transfer (REST) are some standard patterns of design used in making architecture of APIs.

Google Map uses Mercator projection as it tells the shortest possible paths between places. In Mercator, cylindrical projection is used which draws an imaginary tangent line along the equator to give constant directions and thus it shows the shortest and best distance between places. However, the areas at poles are greatly distorted thus show larger than their actual

size^[1]. While Other APIs like Microsoft Bing, Foursquare, Open street, Carto DB, Yahoo Boss place finder, Mapbox GL and Mapbox APIs ^[2] use Representational State Transfer (REST) based architecture which helps to geocode large sets of spatial data through Distributed hypermedia ^[14]. REST allows better navigation by geocoding an address, retrieving images, creating static maps with pushpins and creating routes through GIS like Geocoding, reverse Geocoding, routing and Static imagery. REST provides Mapnik, a toolkit for rendering images in these APIs^[12]. The navigation program is calculated in Map quest via Navteq architecture and eventually compared to other feasible routes between the two spots in order to choose the most appropriate route^[3]. Navteq uses GDT (Geographic Data Technology), DMTI Spacial (Digital moving target indicator), and Tele Atlas in order to search data from which is then used to depict the network of roads in a specific geographic area. Open ^{layers} uses map renderer which is basically a camera, this camera is shared by multiple maps and consists of layers list. The layers have multiple supported projections and each layer publicizes its supported projection. The layer renderer options object stores view-related states for a particular layer and include features of bus, capacity, <http://features.openlayers.org/>, ISSN 1947-5500

visibility, saturation, etc. Options of the layer renderer object have a layer. Multiple layer renderer options may share the same layer. Therefore, a layer can be out looked in unlike fashions. The map renderer communicates to events and creates layer renderers. There is a control that can listen to map events and camera events. The map navigation control acts on the camera. Its types can be described in MVC (Model view controller) terms^[4].



Mercator projection in Google Map ^[1]

1.3 Technology used

Google map, Microsoft Bing, Map quest, Open layers, Open street, Mapbox, Carto db, Yahoo boss place finder map APIs use JavaScript. Google map and Microsoft Bing also use AJAX and XML for web development. Ruby on rails and XML is preferred by Open Street while designing web. In addition to this Map quest uses Curl technology, Open layers makes its work in HTML 5, CSS, ecmaSCRIPT 5. Foursquare have Pilgrim/ Swarm in terms of technology and uses php/scala when it comes to programming. Mapbox GL prefers Open GL 2.0 on web and C++ while programming. Google Map is comfortable in C++ when it comes to programming. Microsoft Bing uses Typescript, C#, Interactive JSON and ASP. Net while Map quest is multilingual. KML and GML find its programming work in Open layers. Open street and Carto db map API relies on Ruby while programming while Mapbox uses Node.js. Python, SDK, JSON. Esri arc GIS uses C++, C#, Visual basic, .NET, Python, java. Yahoo boss place finder makes use of C#, HTML and JSON on web.

1.4 Databases used in various APIs

In Google maps data is spread over vast region, therefore we need a no SQL database which will be able to manage such unstructured data.

Bigtable is a solution for such kind of data. Big table has no limitation of data size and is highly flexible for such kind of dynamic data^[5]. Microsoft Bing on the other hand uses its own product Entity framework 5(dot net framework) to manage its data. Entity framework 5 supports domain specific data and not where the data is actually being stored in the table. Map quest uses Open Database connectivity, an open guideline API to access a database management system. There is no need for a database server as this map API hosts the location data. There is a provision of uploading new location data which can be further modified with the help of web application suite Map quest fast update^{TM [13]}. Open Street, Open layers, Mapbox, Carto db Map APIs use PostgreSQL Also known as Postgres, an open geospatial consortium acquiescent and complete compliance of ACID^[10]. Post GIS used by Mapbox GL is an extension of database for PostgreSQL, an object relational database. Moreover, foursquare map API prefers Mongo DB(a no SQL database)^[9] as this provides auto sharding layers facility helping to scale over many nodes which enables foursquare to manage its data spread over many nodes^[6]. ALTI database used by Esri arc GIS has a unique memory feature which no other database supports^[8]. It has a mixed structure that enables tables residing in the disk with the tables in the memory making use of a sole interface. SAP HANA is also used by Esri arc GIS, a relational database management system^[11], facilitates ETL proficiency and inquisitive data processing apart from being an application server. Yahoo boss place finder uses its own product Yahoo Query Language (YQL) for database. This lets developers run apps faster with very few lines of code and relatively much smaller network footprint. YQL helps developers' access and shape data across internet through one simple language eliminating the need to learn how to call different APIs^[7].

1.5 Accessible platforms

Google map, Map quest, Mapbox, Foursquare, Open Street, Mapbox, Carto DB, Esri arc GIS and Mapbox GL map APIs are supported by android while Microsoft Bing map AP supports Windows devices and foursquare both android and Windows.

Most of the mentioned APIs are open source except a few like Google MAP, Microsoft Bing, Yahoo Boss place finder.

1.6 Comparative analysis of Different Map APIs

| Map | Architecture | Technology | Database Technology | Open Source | Platform | Language | Support Android |
|-----------------------|--|--|---------------------------|-------------|------------------------------|---------------------|-----------------|
| Google map | -Mercator projection | -JavaScript - XML -Ajax | -Big Table | No | -Android -ios -Web | -C++ | Yes |
| Microsoft Bing | -REST based : Geocoding, reverse Geocoding, routing and Static imagery | -JavaScript -AJAX -Typescript -Interactive SDK –JSON -XML .NET | -Entity framework 5 | No | -Windows devices - Web | -ASP .NET -C# | No |
| Map Quest | -Navteq -GDT -DMTI Spacial -AND data solutions -Tele Atlas | -Curl - JavaScript | -ODBC | Yes | -ios -Android | -MultiLingual | Yes |
| Open layers | -Tile grid, -map renderer, -camera layer and -control MVC | -javascript -HTML5 - CSS - ecmaSCRIPT 5 -KML (Keyhole markup language) -GML (Geography markup language) | -Post GIS | Yes | -Web | Node.js | No |
| Foursquare | - REST based : Collects and organises geo tagged data | - Pilgrim/swarm | -Mongo DB | Yes | -ios -Windows -Android | -php/Scala | Yes |
| Open street | - REST based : rendering tiles -rendering maps as raster image - mapnik | -XML -javascript -Ruby on Ralis | -Postgre SQL | Yes | -Web - Android | -Ruby | Yes |
| Mapbox | - REST based : Imagery tiles from efficient compression. It has fast compositing and resilient architecture | -Python SDK -JSON -javascript | -PostgreSQL - Post GIS | Yes | -Web - ios - Android | -Python -Node.js | Yes |

| | | | | | | | |
|------------------------------------|--|---------------------------------------|--|-----|--------------------|--|-----|
| Carto db | -REST based -mapnik -cartocss -leaflet -MBTiles | -JavaScript -Ruby on Rails | -Post GIS -Postgre SQL | Yes | -Android -Web | -Ruby | Yes |
| Esriarc GIS | -2D&3D vector tile byers renderess tasks geometry symbology popups & navigation | -.NET | -ALTIBASE -IBM DB2 -IBM -Informix Microsoft - SQL Server -Oracle -PostgreSQL - SAP HANA -Teradata database | Yes | - Android - Web | -C++ -C# -Python -java -Visual basic | Yes |
| Yahoo boss place finder | REST based : - customization of I frame | -JSON -JavaScript -HTML .NET | Yahoo query language | No | -Web | -C# | No |
| Mapbox GL | REST based : - Vector Tile Format | -Open GL ES 2.0 | Post GIS | Yes | -ios -Android | -C++ | Yes |

1.7 Conclusion

The study compares different Map API's and highlights the differences in their architecture, technology, programming language, database, platform, languages preferred in various APIs. Finally, this review shall study the prospects of mentioned APIs and suggest one to take advantage of particular API's functionalities.

References

- [1] www.GISGeography.com
- [2] www.programmableweb.com
- [3] Mike Milici & Xue Tan
- [4] <http://tschaub.net/openlayers.github.io/en/master/apidoc/tutorial-ol3.html>
- [5] www.ripublication.com/ijcir17/ijcirv13n8_22.pdf (IJCIR [Shagufta Praveen and Dr. Umesh Chandra])
- [6] www.mongodb.com
- [7] www.npmjs.com
- [8] <https://en.wikipedia.org/wiki/Altibase>
- [9] www.mongodb.com
- [10] <https://www.postgresql.org>
- [11] https://en.wikipedia.org/wiki/SAP_HANA
- [12] https://en.wikipedia.org/wiki/Representational_state_transfer
- [13] directionsmag.com

[14] msdn.microsoft.com

Survey of Feature Selection / Extraction Methods used in Biomedical Imaging

Misbah Ziaye
CSIT, Department
MUST University (Mirpur, AJK)
misbahch777@gmail.com

Samina Khalid
CSIT, Department
MUST University (Mirpur, AJK)
Samina.csit@must.edu.pk

Yasir Mehmood
CSIT, Department
MUST University (Mirpur, AJK)
Yasir_mehmood@must.edu.com

Abstract: *Feature selection/extraction methods aimed to reduce the Microarray data. Basically in this comparative analysis, we have taken into account different feature selection and extraction strategies used up till now in the field of Biomedical. In the field of pattern recognition and biomedical imaging, dimensionality reduction is the central area of the research. Some mostly used features selection/extraction methods aim to analyze the most efficient data and achieve the stable performance of the algorithms, as well as improve the accuracy and performance of the classifier. This analysis also highlights widely used dimensionality reduction techniques used up till now in the field of biomedical imaging for the purpose to explore their potency, and weak points.*

Keywords- *Feature Selection, Feature Extraction, Relief, CBM, PCA, GA, MRMR, ICA.*

I. INTRODUCTION

Feature selection methods are widely used for selecting the most relevant and useful features from large datasets. The main difference between features selection / extraction processes is that feature selection method are used to achieve the subset of the most related features without repeating them. Feature extraction methods are used to decrease dimensionality by combining existing features [2]. A feature selection method can improve knowledge of the process and accuracy of learning algorithms as well as its need is considered in data mining and machine learning applications. These feature selection methods have recently been used in the field of biomedical imaging for the automated diagnosis of lung cancer, breast cancer and tumor etc. According to Sansui Cyber Based Image Retrieval (CBIR) techniques are widely used in biomedical images for feature selection/ extraction processes. Feature selection gain the important information from existing features and achieve the highest accuracy of classifiers [3].

Feature selection methods also find the subset of the original and most relevant features into the vector for efficient computations. There are different ways to reducing the dimensionality of microarray data of cancer patients. Large amount of data to be analyzed through dimensionality reduction methods is essential in order to get the meaningful results. In this paper different feature selection and extraction methods are discussed and a comparative analysis is also presented [4]. Several feature selection and extraction methods are used for the purpose of increasing the accuracy of classifiers and reducing the computational complexit. There are vrious automated diagnostic systems have been developed using machines learning and pattern recognition techniques with the combination of image processing techniques. To achieve good accuracy of classification systems and to reduce computational time use of feature selection and extraction methods is really beneficial [5]. Researchers have also acknowledged various practical complexities related to feature selection\extraction methods when used in the field of biomedical imaging. These complexities are also highlighted in this paper. Some of the feature extraction methods like Principle Component Analysis (PCA), which is suitable for better noise tolerance and to avoid over –fitting problems can be efficiently used to achieve good accuracy of classification system used in automated diagnostic systems. Studies show that a lot of research have been done for the suitable feature selection\extraction method and ,worth mentioning approaches are mRmR, CMIM, correlation coefficient, BW-ratio, INTERACT, GA, SVM-REF, RELIEF, (principle component analysis) , (Non-Linear) principal component analysis, (independent component analysis), and correlation based feature selection. . A brief survey of these techniques based on literature review is performed to check suitability of different feature selection\extraction techniques in certain situation based on experiments that have been performed by researchers to analyze how these techniques helps to improve predictive accuracy of classification algorithm.

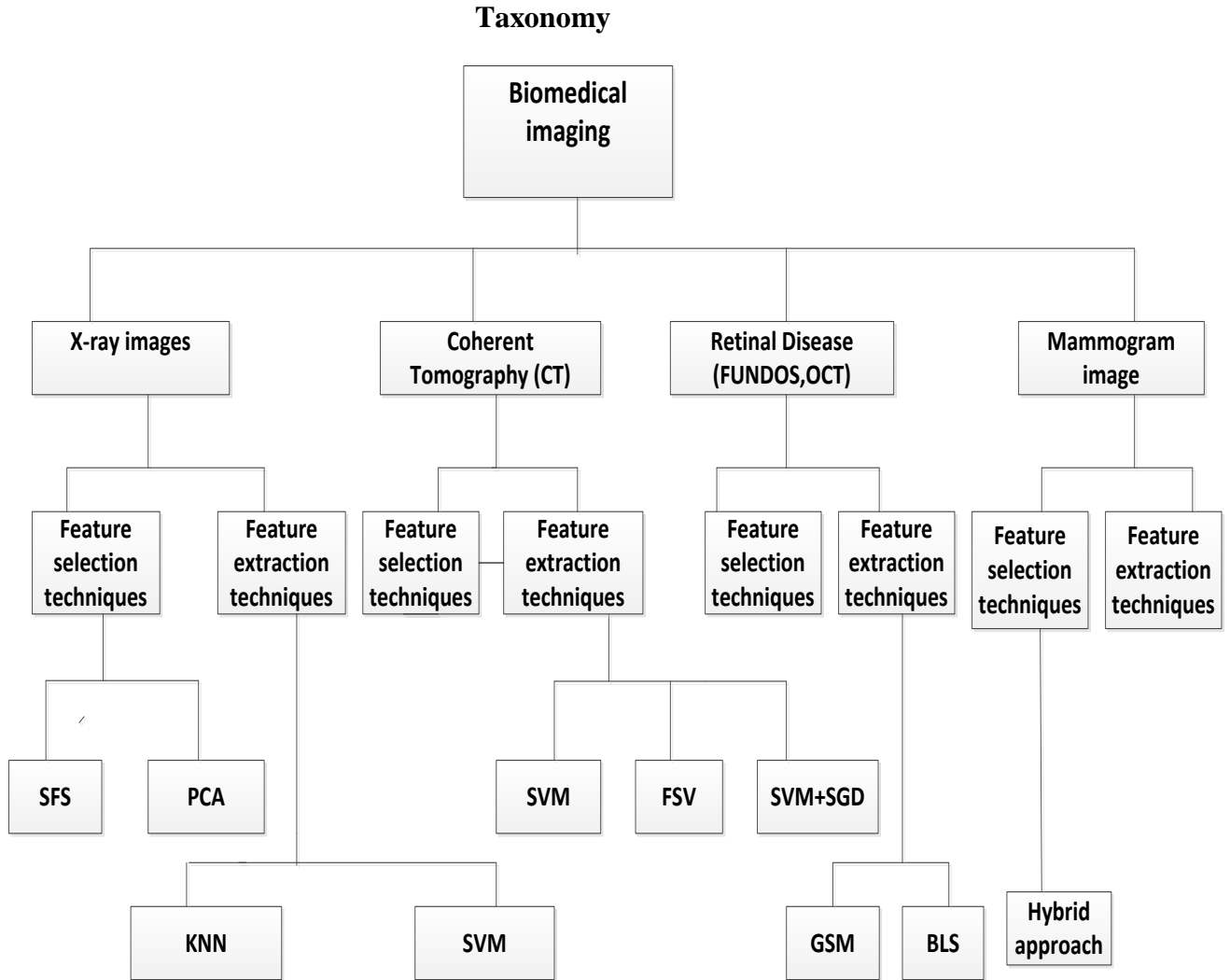


Fig. 1 Taxonomy

Feature selection\ extraction methods that have been used up till now in the field of biomedical imaging are explained in next section. These methods are categorized from the perspective of automated diagnostic system developed for different diseases as can be seen in figure 1. A taxonomy is presented in this paper showing different level of our research analysis performed in this study, in taxonomy biomedical imaging is presented as a root node and then in next level these images are classified according to diseases for which these images are analyzed. In next level the emphasis of this study is to analyze which feature selection/extraction methods are specifically used further in classification systems, proposed for the automated diagnostic system for specific disease.

II. FEATURE SELECTION AND FEATURE EXTRACTION METHODS USED IN BIOMEDICAL IMAGING

There are several feature selection\extraction methods used in biomedical imaging for increasing the amount of accuracy as well as reducing the computational complexity of existing methods. Different feature selection/ extraction methods were described and compared in next section.

A. Esophageal cancer detection from X-ray images:

Esophageal cancer is very common now a days. In [1] an automated diagnostic system developed for timely detection of this disease. In this automated systems feature selection and extraction methods PCA (principal component analysis),SVM (support vector machine) is used to diagnosis the disease of esophageal cancer. Due to limitations of the current studies there is need of more work

on Esophageal cancer and also improved the diagnostic quality of disease and more advanced feature selection/extraction methods, are used for segmentation of esophageal X-ray images as shown in figure 2.

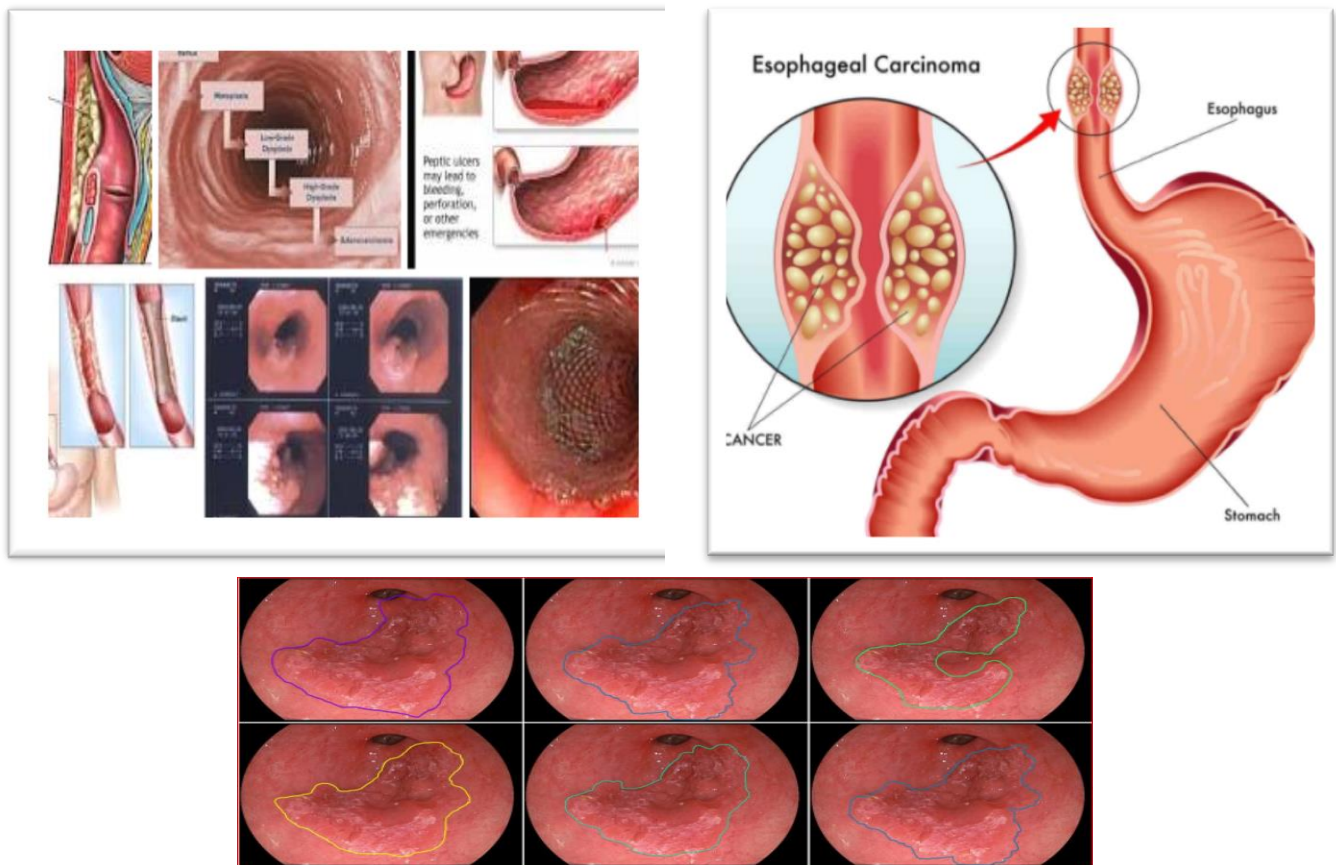


Fig 2. X-ray images of esophageal cancer [2]

B. ARMD Disease Detection Using Fundus, OCT Images

Feature selection\extraction techniques used in many ophthalmologists diseases. In [9] Feature Extraction method is more suitable for automated detection of ophthalmology disease. Now here Probabilistic Neural Network (PNN), method is presented for OCT images.

Figure 3 (a) Shows in this paper healthy drusen detection Figure 3 (b) illustrates the hard drusen, Figure 3 (c) shown the soft drusen respectively. Fig.3 presented some methods for feature extraction KNN (K nearest neighbour), (Probabilistic Neural Network), and (independent principal component analysis) is used for these categories of images based on (clustering), (edge detection), or (thresholding), as well as (template matching)

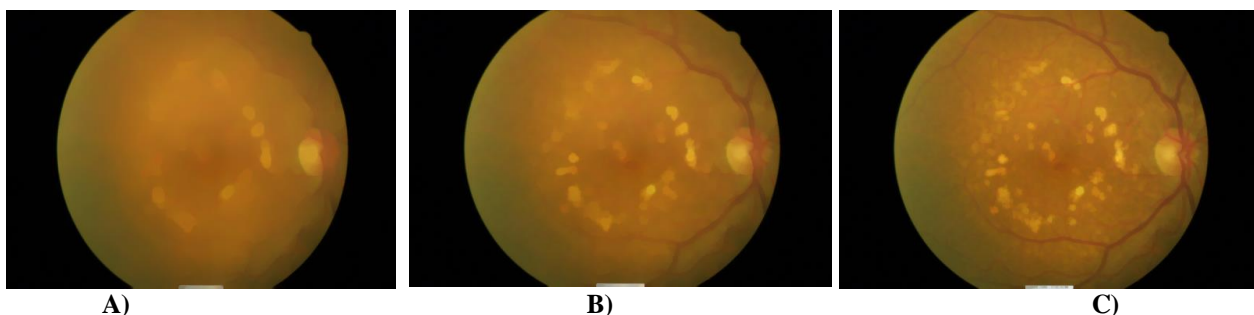


Fig 3 automated drusen detection Images. [10]

C. Lung Cancer detection using CT scans images:

Most common types of tumors, with the highest mortality, and morbidity. The Radionics features, and

Pretherapy of CT images are used to predict the distant and metastasis images [7].

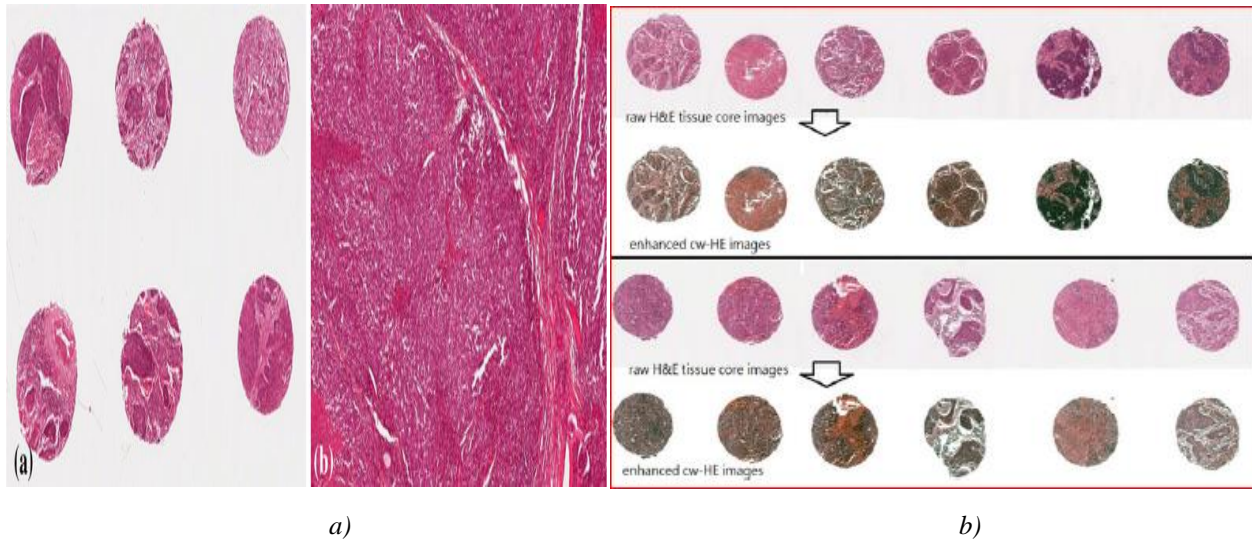


Fig.4 Lung Cancer detection [12]

Fig.4 (a) presented the Tissue samples shape and Fig.4 (b) showing the Tissue section It can also use the tumor, phase information.

Medical imaging and distant meta states modalities are also used in lung cancer. Lung cancer treatment is more difficult surgery, almost impossible in many cases to extract the cancer calls. Feature selection method GLR (gray-level run) is use to solve many critical problems related to lung cancer.

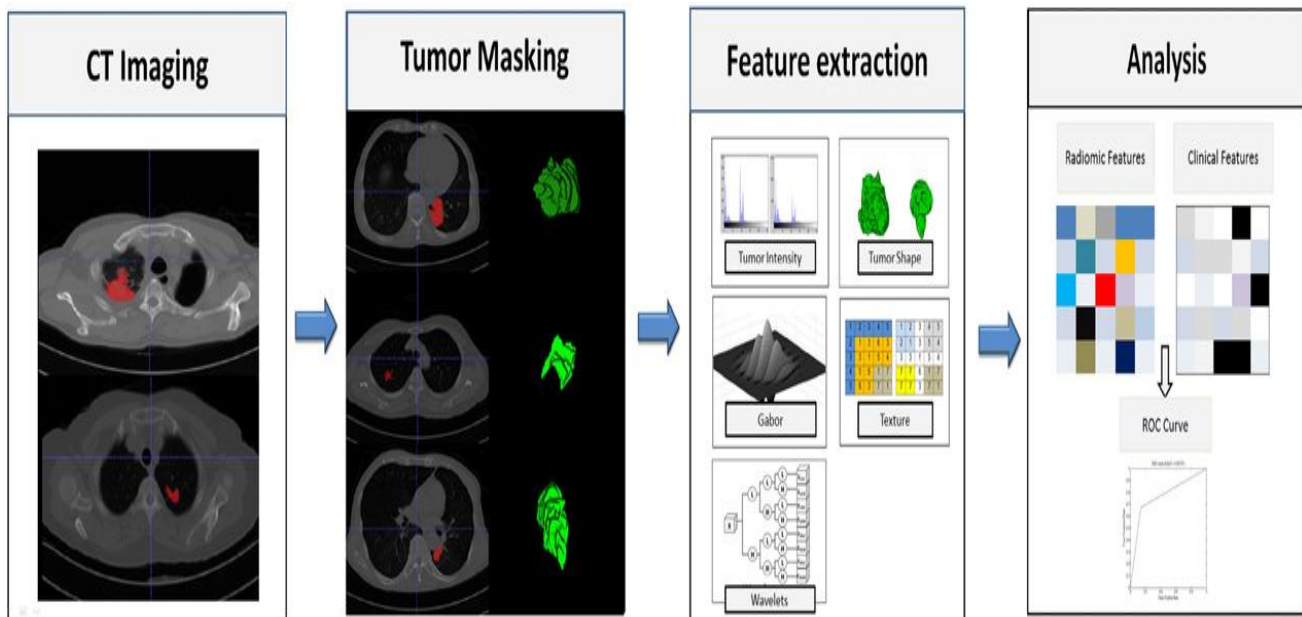


Fig 5 CT images of lung cancer [4]

The following Fig 5 presented the strategy of radiomic data and extracting the tumor masking features from CT images. Red section in fig 5 illustrate the tumor area.

Basically features are extracted based on shape and texture. Radiometric features are combined for analysis in clinical data.

D. Breast cancer detection using mammogram images

Most common forms of breast cancer are found in the Womens. In India 1 to 22 women, are suffering due to breast cancer. The calcification of breast tissues and accurate detection shown in Fig 6. Basically, noise reduction applied in a preprocessing step to improve the classification contrast and image results. The noise reduction step separated the malignant images and normal images [16]. The image classifier is used to classify the mammogram images and mammogram image, is classified the normal images.

Fig. 6 a) illustrates the ROI start processing and separated the normal and disease images. Fig 6 (b) shows ROI after Pre-processing Operation. By comparing the image (a) and image (b) we observe the background mammography noisy features are removed. The hybrid approaches for the Feature selection, is proposed which can also reduce 75 percent of the features, Into original images. The decision tree algorithms, can apply in mammography classification which can also used to reduce the large and noisy features.

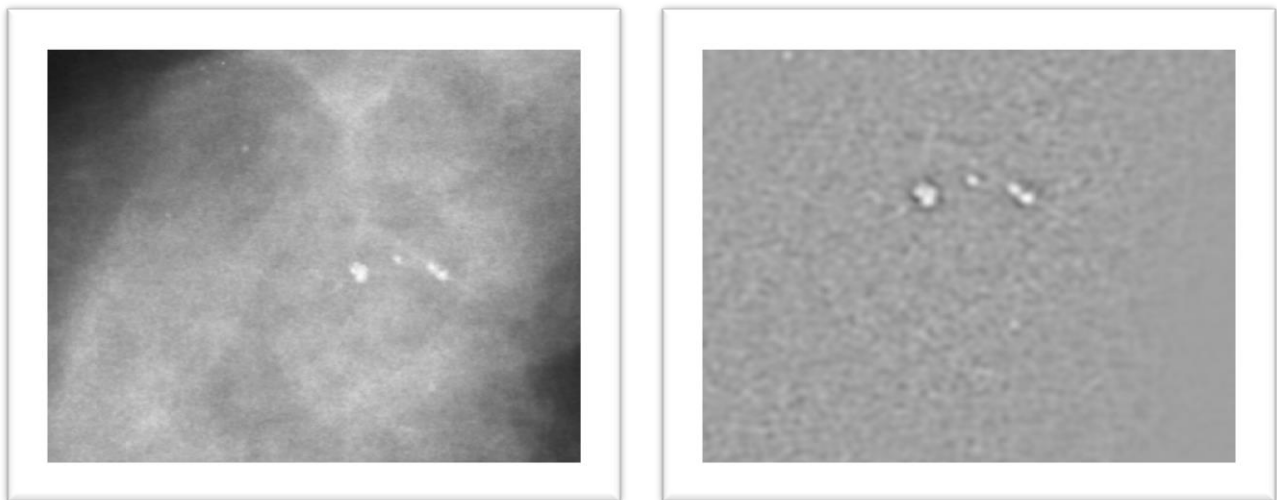


Fig. 6 Mammogram Images of breast cancer [3]

LITERATURE REVIEW

Hira, Z.M. and D.F. Gillies, et al [3] presented the Cybir Based Image Retrieval (CBIR) techniques are widely used in biomedical images for feature selection/ extraction processes. (CBIR) Central area of research in medical above the decade of the 10 years. Feature selection gain the important information from existing features and achieve the highest accuracy of classifiers. Feature selection methods also find the subset of the original and most relevant features into the vector for efficient computations. Saeyes, Y., I. Inza et al [4] introduced the concept of automated diagnostic systems. There are various automated

diagnostic systems have been developed using machine learning and pattern recognition techniques with the combination of image processing techniques. To achieve good accuracy of classification systems and to reduce computational time use of feature selection and extraction methods is really beneficial.

Kunasekaran, et al [5] proposed some of the feature extraction methods like Principle Component Analysis (PCA), which is suitable for better noise tolerance and to avoid over-fitting problems can be efficiently used to

achieve good accuracy of classification system used in automated diagnostic systems.

Zhou, H., et al [6] discussed the analyses of features selection and extraction methods used in biomedical image processing. Basically, how images can choose through diverse site. A feature selection method can improve knowledge of the process and accuracy of learning algorithms as well as its need is considered in data mining and machine learning applications. These feature selection methods have recently been used in the field of biomedical imaging of lung cancer.

Wosiak, A et al [7] proposed the (ACO), Ant Colony Optimization for feature extraction is mainly used in biomedical imaging. Increases the complexity of the system when a large number of features extracted from the image. Ant Colony Optimization extracting the most useful and relevant features from the image and also improve the system accuracy and decreases the system complexity. ACO approach is less expensive and give a good results as compare to other methods.

Samina Khalid, Tehmina Khalil et al [8] introduced some mostly used features Selection / extraction methods with the basic purpose to explore and analyze how efficiently of such methods can used to achieve best performance, for learning algorithms and as well as improve the accuracy of the predictive classifiers and learning algorithms. The dimensionality reduction method can, briefly to explore the potential of a weak point. And mostly used in dimensionality reduction methods to achieve high accuracy. Gnanasekar, et al [9] proposed the SVM.RELEF methods for feature selection. These methods select the feature and also improving the accuracy of classifiers. Some algorithms can be extracted new feature and selected by the existing features. These methods can be used in diagnosis disease as well as achieve the endoscopic, surgery system. The proposed methods can also extract two types of features.

In [10] Feature selection methods are widely used for selecting the most relevant and useful features from large datasets. The main difference between features selection / extraction processes is that feature selection method are used to achieve the subset of the most related features without repeating them. Feature extraction methods are used to decrease dimensionality by combining existing features. A feature selection method can improve knowledge of the process and accuracy of learning algorithms as well as its need is considered in data mining and machine learning applications.

Soliz, P., et al et al [11] presented different ways of reducing the dimensionality of microarray, cancer data. In this paper different feature selection and extraction methods are discussed and a comparative analysis is also presented. Several feature selection and extraction methods are used for the purpose of increasing the accuracy of classifiers and reducing the computational complexity. There are various automated diagnostic systems have been developed using

machines learning and pattern recognition techniques with the combination of image processing techniques.

Fu, D., et al [12] proposed the dimensionality reduction of Feature selection / extraction techniques is presented to detect many ophthalmologists disease In this paper GSM, BLS methods for feature extraction is presented for ARMD disease.

Belle, A. et al [13] Proposed technique that extracts the features for classification. GSCM (gray scale component matrix) method can apply in retinal disease. And as well as 100 images of dataset is used to detect the disease and normal images. Image can be grouped into different classes base on the VC visual distinctiveness. Independent components analysis (ICA) Technique is used to detect the features into input classifiers. Independent components analysis (ICA) Technique also extract the basic and important features into phonotype images.

Hira, Z.M. et al [3] introduced the concept of diabetic detection using feature extraction method. Now here that artificial neural networks and (PCA) principle component analysis is presented for diabetic disease. The basic results shown artificial neural networks singular value of decomposition+ (PCA) principle component analysis, is most suitable for diabetes disease detection. These methods achieve highest accuracy with less cost as well as minimum computation time.

In [14] are presented the Co focal Scanning Laser Tomography (CSLT) approach, that can be involve to analysis the tumor disease images. Moment method for feature selection is used to improves the accuracy of the predictive classifiers and learning algorithms. GA, KNN methods for feature selection is used to detect the automated and semi-automated tissues for lung cancer. Basically multiple transformations of different preprocessing steps can be perform to extract the necessary and relevant data from the image.

Boughattas et al [15] proposed method of different images that can be choosed from input subset and photographic images are primarily because they can relevant and internal anatomy, features are same. Sanusi said these techniques are proposed for feature selection / extraction classification and retrieval, (CBIR) is widely used in biomedical images. Information Gain and achieve the structure of the features set and also find the subset of the original and most related features into the vector and efficient computation.

In [16] proposed the Most common form of breast cancer is found in the womens. In India 1 to 22 women, is suffering due to breast cancer. The Hybrid approach for mammogram images is classified into the normal images and process start the malignant image are separated. Total 26 Features were included in (histogram intensity) and Features of GLCM, are also Extracted into mammogram images. The hybrid approaches for the Feature selection, is proposed which can also reduce 75 percent of the features, into original images. The decision tree algorithms, can also apply in

mammography classification which can also use to reduce the large and noisy features.

George, et al [17] presented the feature selection and extraction methods used to extract the genes marker (GM) that can be influenced on the classification accuracy as well as effectively, eliminating the redundant, noisy and repeated features. They proposed some feature selection techniques can be used for cancer classification. DSS (decision support systems) method for feature extraction also applied in genes marker (GM) to extract the suitable features.

In [18] Most common types of tumors, with the highest mortality, and morbidity. The Radionics features, and pretherapy of CT images are used to predict the distant metastases images. In this paper SFS and PCA methods for feature selection is used to detect the tissue sample. Feature extraction methods SVM, KNN is used for tissue samples and also extract the lung cancer images. The results show SVM, KNN give most accurate results as compare to feature selection methods.

III DISCUSSION

The dimensionality reduction method can, briefly to explore the potential of a weak point (what is this statement about?). And mostly used in dimensionality reduction methods to achieve high accuracy. Feature extraction PSA (using principle component analysis), methods suitable for better noise tolerate, procedures. There are different ways to reducing the dimensionality of microarray data of cancer patients. Large amount of data to be analyzed through

dimensionality reduction methods is essential in order to get the meaningful results. In this paper different feature selection and extraction methods are discussed and a comparative analysis is also presented. Several feature selection and extraction methods are used for the purpose of increasing the accuracy of classifiers and reducing the computational complexity. There are various automated diagnostic systems have been developed using machine learning and pattern recognition techniques with the combination of image processing techniques. To achieve good accuracy of classification systems and to reduce computational time use of feature selection and extraction methods is really beneficial. Esophageal cancer is more common and also used in the diagnosis of disease. Feature extraction/selection techniques are used for segmentation and esophageal of X-ray images. Feature selection/extraction techniques used in many ophthalmologists disease also. Now here GSM, BLS methods for feature extraction is presented for ARMD disease. The most common form of breast cancer is found in women. Due to breast cancer 1 in 22 women in India is likely, suffer due to breast cancer. Mammogram images are classified into the normal images and process start the malignant image are separated. Hybrid approach for feature selection, is presented which can also reduce the 75% of the features into original images. SVM and RFE feature selection, technique that can be combined with the SMO classifier has been demonstrating its potential ability as well as accurately and efficiently work on classifying both binary and multiclass high dimensional sets of the tumor specimens.

Table: features selection and extraction methods

| Ref | Problem | Proposed Methods | Techniques | Result |
|-----|---|------------------------------------|--|---|
| 1 | High dimensional data | Dimensionality reduction | MRMR, PSA | MRMR technique works better than all other techniques |
| 2 | Irrelevant/, redundant features. | Dimensionality reduction technique | (CSLT), GA, hybrid | |
| 3 | (Biochemistry) and (medicine) is Important problematical: | GA/knn | Non hypothyroid, hypothyroid | The masking (GA/knn) achieved best accuracy |
| 4 | Dimensionality problem. | SVM | DNA MICROARRAY | SVMs show better performance. |
| 5 | Shape modeling problem | Cortical folding | Post processing, reconstruction techniques | |

| | | | | |
|----|--|-------------------------------|----------------------------------|---|
| 6 | Tumor Motion tracking. | XCAT phantom, | CBM, PSA and Relief | CBM +GA proved the best performance accuracy. |
| 7 | (Redundant) features. | | (CBIR), (ACO), (CSLT),GA,, (CAD) | |
| 8 | Difficult to Extract the cancer features | PCA, NIR, Hyper spectral Data | CNN | CNN model achieved the 75.9% accuracy |
| 9 | Diagnosis of Glaucoma. | Computer aided analysis | (RNFL), Knn,LM,, SFFS, (LCP) | LM achieves, highest accuracy |
| 10 | JRC problem | Relational Regularization | (ADAS Cog, MMSE) | MMSE gives the best results |

IV OPEN PROBLEMS

- Due to limitations of the current studies there is need of more work on Esophageal cancer and also improved the diagnostic quality of disease and more advanced feature selection/extraction methods, are used for segmentation of esophageal X-ray images.
- Also improved the diagnostic quality system more advanced feature extraction and selection methods are used for segmentation and X-ray images.
- Further improve the performance of (CT) images and phenotypes, Features are necessary.
- How we improve the processing methods can also enhance visual interpretation as well as image analysis. And also provide automated or semi automated tissue detections.

V CONCLUSION

Classification of the high dimensional biomedical data sets is the most difficult task..The main difference in features selection / extraction feature selection method are used to achieve the subset of the most related features without repeating them . Feature extraction methods are used to decrease dimensionality by combining existing features. The feature extraction method uses to decrease dimensionality

reduction and already existing features are produced that can be more relevant and significant.A feature selection methods improve, knowledge of the process and accuracy of learning algorithms as well as its need is considered in data mining and machine learning applications some of widely used feature selection techniques in lung cancer, breast cancer tumor and etc. Due to limitations of the current study also need for more work on Esophageal cancer as well as Breast, lung, and retinal diseases. Also improved the diagnostic quality more advanced feature extraction and selection methods are used for segmentation and X-ray images. Further improve the performance of (CT) images and phenotypes, Features are necessary.

REFERENCES

1. Yang, F., et al., *Feature Extraction and Classification on Esophageal X-Ray Images of Xinjiang Kazak Nationality*. Journal of healthcare engineering, 2017. **2017**.
2. Pujari, P. and J.B. Gupta, *Improving classification accuracy by using feature selection and ensemble model*. International Journal of Soft Computing and Engineering, 2012. **2**(2): p. 380-386.
3. Hira, Z.M. and D.F. Gillies, *A review of feature selection and feature extraction*

- methods applied on microarray data.* Advances in bioinformatics, 2015. **2015**.
4. Saeys, Y., I. Inza, and P. Larrañaga, *A review of feature selection techniques in bioinformatics.* bioinformatics, 2007. **23**(19): p. 2507-2517.
5. Kunasekaran, K. and R. Sugumaran. *Exploratory analysis of feature selection techniques in medical image processing.* in *Proceedings of the international conference on information engineering, management and security.* 2016.
6. Zhou, H., et al., *Diagnosis of Distant Metastasis of Lung Cancer: Based on Clinical and Radiomic Features.* Translational oncology, 2018. **11**(1): p. 31-36.
7. Wosiak, A. and A. Dziomdziora, *Feature Selection and Classification Pairwise Combinations for High-dimensional Tumour Biomedical Datasets.* Schedae Informaticae, 2015. **24**: p. 53.
8. Khalid, S., T. Khalil, and S. Nasreen. *A survey of feature selection and feature extraction techniques in machine learning.* in *Science and Information Conference (SAI), 2014.* 2014. IEEE.
9. Gnanasekar, P., et al., *Investigation on feature extraction and classification of Medical images.* World Academy of Science, Engineering and Technology, 2011. **60**: p. 327-332.
10. Rangarajan, L., *Bi-level dimensionality reduction methods using feature selection and feature extraction.* International Journal of Computer Applications IJCA, 2010. **4**(2): p. 33-38.
11. Soliz, P., et al. *Independent component analysis for vision-inspired classification of retinal images with age-related macular degeneration.* in *Image Analysis and Interpretation, 2008. SSIAI 2008. IEEE Southwest Symposium on.* 2008. IEEE.
12. Fu, D., et al., *Retinal status analysis method based on feature extraction and quantitative grading in oct images.* Biomedical engineering online, 2016. **15**(1): p. 87.
13. Belle, A., M.A. Kon, and K. Najarian, *Biomedical informatics for computer-aided decision support systems: a survey.* The Scientific World Journal, 2013. **2013**.
14. Nankali, S., et al., *Optimum location of external markers using feature selection algorithms for real -time tumor tracking in external -beam radiotherapy: a virtual phantom study.* Journal of applied clinical medical physics, 2016. **17**(1): p. 221-233.
15. Boughattas, N., et al. *Brain tumor segmentation from multiple MRI sequences using multiple kernel learning.* in *Image Processing (ICIP), 2014 IEEE International Conference on.* 2014. IEEE.
16. Vasantha, M., V.S. Bharathi, and R. Dhamodharan, *Medical image feature, extraction, selection and classification.* International Journal of Engineering Science and Technology, 2010. **2**(6): p. 2071-2076.
17. George, G. and V.C. Raj, *Review on feature selection techniques and the impact of SVM for cancer classification using gene expression profile.* arXiv preprint arXiv:1109.1062, 2011.
18. Magdy, E., N. Zayed, and M. Fakhr, *Automatic classification of normal and cancer lung CT images using multiscale AM-FM features.* Journal of Biomedical Imaging, 2015. **2015**: p. 11.

A Novel for Finding Missing Piece in the Borderless Square Jigsaw Puzzle

Sendi Novianto, Luo Fei

School of Automation, South China University of Technology, Guangzhou – Guangdong, China, 510640

sendi.novianto@dsn.dinus.ac.id, aufeiluo@scut.edu.cn

Abstract - There are many object detection methods have been proposed for detecting some objects or colors in images. The jigsaw puzzle pieces of a piece become a challenge in the artificial intellegent world, this challenge becomes more difficult when there are some pieces missing, so it is certain that the formation of the jigsaw puzzle image becomes incomplete. This paper presents a novel for finding missing piece in the borderless square jigsaw puzzle using square detection method with Blob Analysis combined with genetic algorithm using dynamic parameter, which is capable to detect missing pieces in square jigsaw puzzle image. Our method has worked by skipping pixel by pixel detection in each square. This results in a fast timing piece detection in square jigsaw puzzle image. The examples of the result of the application that use our method to detect missing pieces in jigsaw puzzle are also shown. The detection of missing pieces has already done with 25% until 99% of missing pieces.

I. INTRODUCTION

Shape detection is needed in many computer vision tasks because shape is an important clue for modelling objects in scenes. Object location problems are mainly solved by two types of techniques: In one hand, deterministic techniques include application of Hough transform based on methods[1][2][3], geometric hashing[4][5][6][7][8] and template or model matching techniques. In the other hand, stochastic techniques [9][10][11][12] include random sample consensus techniques [13][14][15], simulated annealing [16][17] and genetic algorithms (GA) [18][19][20][21][22][23].

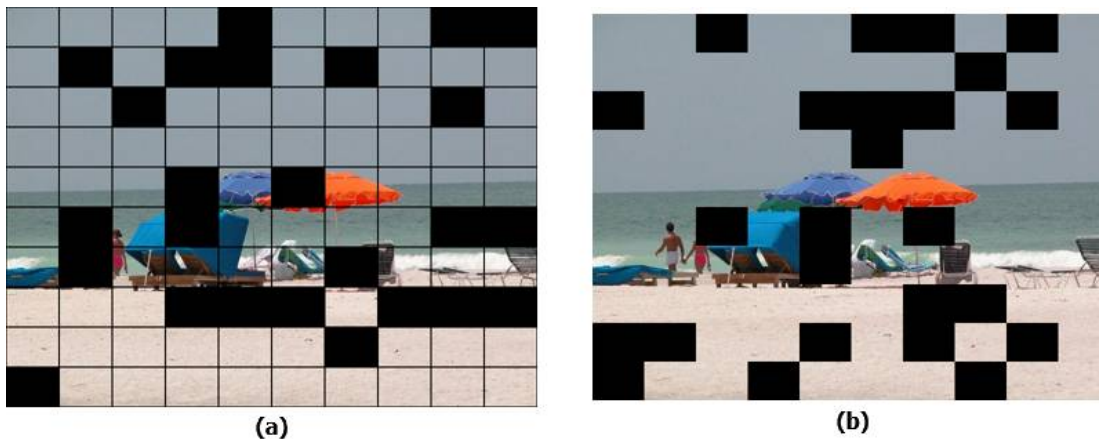


Figure 1. Square Jigsaw Puzzle. (a) with border and (b) without border

Jigsaw puzzle is one of the brain games. To solve this game, we must put each piece in order to get full image. This game will not finish if there is one missing piece in it. There are some methods that we can use to find missing pieces in the jigsaw puzzle with border[24] or without border[25], more detail can be seen in Figure 1. All this

method already proven to find and labeling missing piece of jigsaw puzzle. In the process of detecting missing pieces, there is still having some problems, much bigger the image of square jigsaw puzzle and more pieces are missing, they will take more time to detect and label it.

Genetic Algorithm is a computational algorithm that is inspired by the theory of evolution which is then adopted into a computational algorithm to find solutions to a problem in a more “nature” way. One application of genetic algorithm is on the combination optimization problem, which is to get an optimal solution value to a problem that has many possible solutions[21]. This paper will discuss the implementation of genetic algorithm to find missing pieces in the image of square jigsaw puzzle.

The contribution of this paper is fourfold. First, we propose Blob Analysis with Genetic Algorithm to handle missing pieces in the square jigsaw puzzle. Second, we use dynamic parameter in Genetic Algorithm to make evolution more shortened. Third, in the finding process, we also use auto switch system, this system works when it detects the black color more wide depends on the other colors. Finally, we adopt a set of parametric for finding smallest missing square in the square jigsaw puzzle.

This paper is organized as follows. Section 2 presents the step of introduction to BADPIG-ASS methods. Section 3 evaluates the performance of the proposed approaches using previous methods dataset. Section 4 draws the conclusions and perspectives also the possible future works.

II. INTRODUCTION TO BLOB ANALYSIS WITH DYNAMIC PARAMETER IN GENETIC ALGORITHM AND AUTO SWITCH SYSTEM (BADPIG-ASS) METHODS.

A. Random Square with Random Size and Position Generator

Requirement two input parameter, first is jigsaw puzzle image in PNG file type (Im_{Input}), and the second is auto switch function, 1 for enable this function, 0 for disabled. There are 4 outputs for this process, first is Square (Sqx) with location (Sqx_x, Sqx_y) and size of width (Sqx_{Width}) and height (Sqx_{Height}). Second is square jigsaw puzzle with black and white colors also Blob Analysis and morphological reconstruction filter ($Im_{BlackWhite}$). Third is the real image from input parameter (Im_{Real}). The final output is an inverse from second output ($Im_{Inverse}$).

All the image data and jigsaw puzzle generator used in this paper was from previous paper [25]. The image data will be use for filling first parameter Im_{Input} . In this paper, we will try to use both of them. From the first parameter will be directly saved into Im_{Real} , other than that we can get information about width ($Im_{InputWidth}$) and height ($Im_{InputHeight}$). In addition, we will make change for the color from the Im_{Input} , where the color is other than black will be changed into white, so it will produce black and white image in RGB format ($Im_{BlackWhite}$). The next step converts $Im_{BlackWhite}$ to gray image (Im_{Gray}). The next step, if the second parameter equals 1, the system will count total number of white color (255) in Im_{Gray} and it will become as $Im_{GrayWhite}$ with total number of white color in $(Im_{Gray})^C$, it will become as $Im_{GrayImcomplementWhite}$. If $Im_{GrayWhite} \leq Im_{GrayImcomplementWhite}$ then $Im_{BlackWhite} = (Im_{mr})^C$ other than that $Im_{BlackWhite} = Im_{mr}$.

$$f\left(\text{Im}_{\text{GrayWhite}}\right) \begin{cases} \leq \text{Im}_{\text{GrayImcomplementWhite}} & \text{Im}_{\text{BlackWhite}} = (\text{Im}_{\text{mr}})^c \\ > \text{Im}_{\text{GrayImcomplementWhite}} & \text{Im}_{\text{BlackWhite}} = \text{Im}_{\text{mr}} \end{cases} \quad (1)$$

If the parameter equals 0, it will not give any effect. The last $\text{Im}_{\text{BlackWhite}}$, we close all small holes that exist with morphological Reconstruction[24] (Im_{mr}). The final result for second output is $\text{Im}_{\text{BlackWhite}} = \text{Im}_{\text{mr}}$, and for the fourth output is $(\text{Im}_{\text{BlackWhite}})^c$ as $\text{Im}_{\text{Inverse}}$. To produce Sq_x , the first step is try to find the longest length (MaxNumber) between $\text{Im}_{\text{InputWidth}}$ and $\text{Im}_{\text{InputHeight}}$, it will be used to produce 4 numbers between 1 and MaxNumber at random. This random number will be used as Sq_x , Sq_y , Sq_xWidth , Sq_xHeight in ordered. In the process to be Sq_x , Sq_y , Sq_xWidth and Sq_xHeight , there are several steps that must be passed. The steps are as follows:

- 1) If $\text{Sq}_x + \text{Sq}_x\text{Width} - 1 > \text{Im}_{\text{InputWidth}}$ then Sq_x will be changed to some random number from 1 until $\text{Im}_{\text{InputWidth}}$, and Sq_xWidth will be changed to some random number from 1 until $\text{Im}_{\text{InputWidth}}$.

$$f\left(\text{Sq}_x + \text{Sq}_x\text{Width} - 1\right) \begin{cases} > \text{Im}_{\text{InputWidth}} & \begin{aligned} \text{Sq}_x &= \text{random}[1.. \text{Im}_{\text{InputWidth}}] \\ \text{Sq}_x\text{Width} &= \text{random}[1.. \text{Im}_{\text{InputWidth}}] \end{aligned} \\ <= \text{Im}_{\text{InputWidth}} & \begin{aligned} \text{Sq}_x &= \text{Sq}_x \\ \text{Sq}_x\text{Width} &= \text{Sq}_x\text{Width} \end{aligned} \end{cases} \quad (2)$$

- 2) If $\text{Sq}_y + \text{Sq}_x\text{Height} - 1 > \text{Im}_{\text{InputHeight}}$ then Sq_y will be changed to some random numbers from 1 until $\text{Im}_{\text{InputHeight}}$, and Sq_xHeight will be changed to some random numbers from 1 until $\text{Im}_{\text{InputHeight}}$.

$$f\left(\text{Sq}_y + \text{Sq}_x\text{Height} - 1\right) \begin{cases} > \text{Im}_{\text{InputHeight}} & \begin{aligned} \text{Sq}_y &= \text{random}[1.. \text{Im}_{\text{InputHeight}}] \\ \text{Sq}_x\text{Height} &= \text{random}[1.. \text{Im}_{\text{InputHeight}}] \end{aligned} \\ <= \text{Im}_{\text{InputHeight}} & \begin{aligned} \text{Sq}_y &= \text{Sq}_y \\ \text{Sq}_x\text{Height} &= \text{Sq}_x\text{Height} \end{aligned} \end{cases} \quad (3)$$

- 3) After getting fixed number of Sq_x , Sq_y , Sq_xWidth , Sq_xHeight , the next step is to check if inside of the Sq_x has a color other than black, then produces 4 numbers between 1 and MaxNumber at random. This random number will be used as Sq_x , Sq_y , Sq_xWidth , Sq_xHeight in ordered. Then, repeat the step number 1 until inside of Sq_x has black color. The result can be seen in Figure 2.

$$\left. \begin{array}{l} \text{if } (\text{sum}(\text{find}(\text{Sq}_x))) > 0 \end{array} \right\} \begin{array}{l} \text{Sq}_{x_x} = \text{random}[1..\text{Max}_{\text{Number}}][0] \\ \text{Sq}_{x_y} = \text{random}[1..\text{Max}_{\text{Number}}][1] \\ \text{Sq}_{x_{\text{width}}} = \text{random}[1..\text{Max}_{\text{Number}}][2] \\ \text{Sq}_{x_{\text{height}}} = \text{random}[1..\text{Max}_{\text{Number}}][3] \end{array} \quad (4)$$

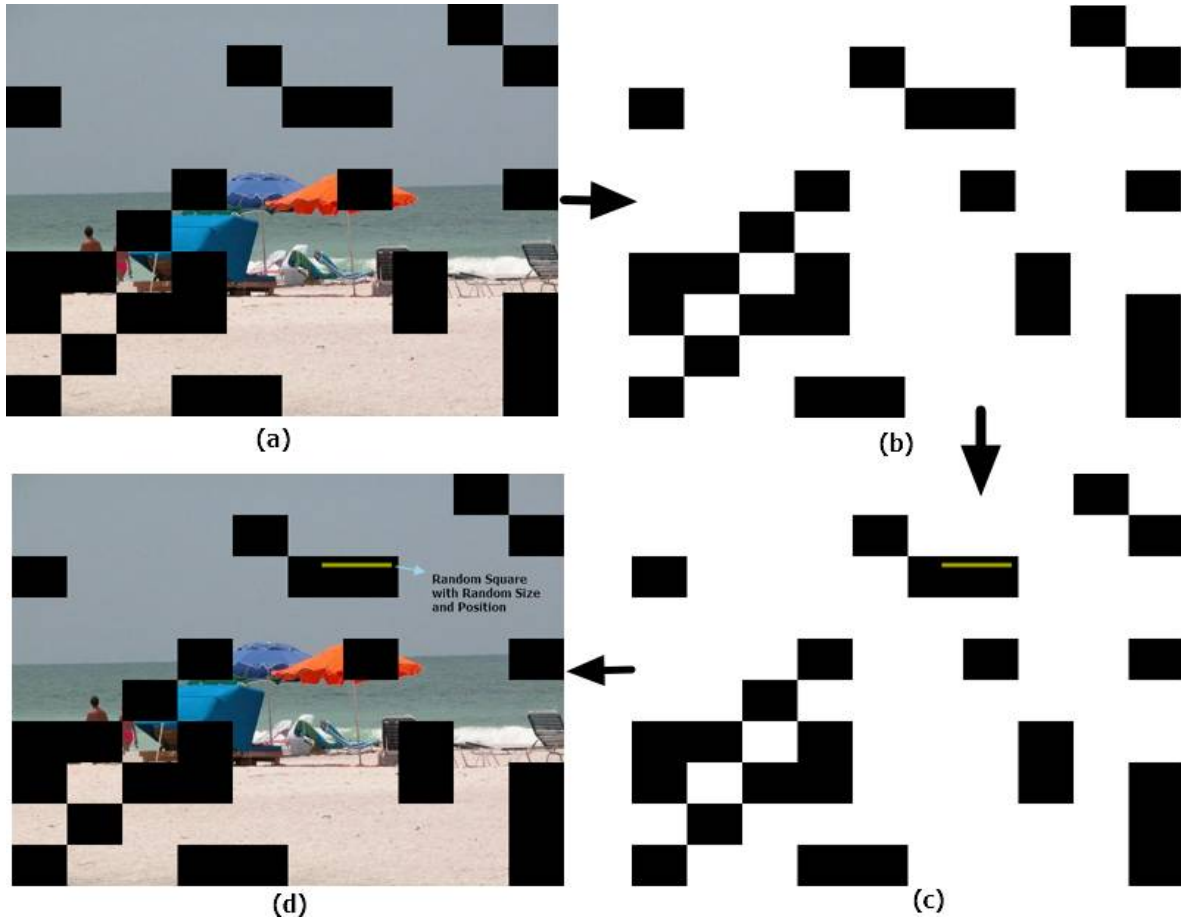


Figure 2. Result of Random Square with Random Size and Position

B. Edge Detection

Based on the result from the first step, all the edge from Sq_x (left, right, bottom and top) will be expanded using GA with dynamic parameter for minimum and maximum number depend on the nearest piece with color. This process requires two input parameters, first is Sq_x and second is $\text{Im}_{\text{BlackWhite}}$. There are two outputs for this process, first is new Square (Sq_N) with new location ($\text{Sq}_{N_x}, \text{Sq}_{N_y}$) and width ($\text{Sq}_{N_{\text{width}}}$) also height ($\text{Sq}_{N_{\text{height}}}$). The second output is used for testing image ($\text{Im}_{\text{tester}}$) to looking maximum or minimum size of edge.

Based on the result above, we get property of Sq_x , with this property we will use GA to manipulate Sq_x . Thus, this Sq_x size will be changed to fill up black area surround them. In our test, we have fixed a number of settings in GA

Table 1. List Of Parameters for The Genetic Algorithm used in the square jigsaw puzzle detection problem

| Parameter | Value |
|------------------|----------------|
| Population Size | 5 |
| Maximal Gene | 100 |
| Number Data | 1 |
| Cross Over rate | 0.25 |
| Mutation Rate | 0.8 |
| Selection Method | Roulette Wheel |
| Crossover Method | One Cut Point |

There are several steps in edge detection with GA, every edge has their own step. This step will start from left, right, top, and bottom of the edge.

1) Initialization and Chromosome Generator

In first step of GA, we need to define Chromosome that we will use. To generate this chromosome, we will get from random number. The random number will depend on which side we will process first, if we want to process:

- left side, the random number for chromosome will be $[1 \dots Sq_{xx}]$
- top side, the random number for chromosome will be $[1 \dots Sq_{xy}]$
- bottom side, the random number for chromosome will be $[Sq_{xHeight} + Sq_{xy} - 1 \dots ImInputHeight]$
- right side, the random number for chromosome will be $[Sq_{xx} + Sq_{xWidth} - 1 \dots ImInputWidth]$

2) Chromosome Evaluation

The problem we want to solve is to find maximum width and height and also minimum position of x and y in Sq_x that are in the black color area. So, the *objective function* that can be used is

- Left side, Sq_{xx} with condition
 - $Sq_{xx} = 1$
 - All color inside the Sq_x is black
- Top side, Sq_{xy} with condition
 - $Sq_{xy} = 1$
 - All color inside the Sq_x is black
- Bottom side, $Sq_{xy} + Sq_{xHeight} - 1$
 - $Sq_{xy} + Sq_{xHeight} - 1 = ImInputHeight$
 - All color inside the Sq_x is black

d. Right side, $Sq_{xx} + Sq_{xWidth} - 1$

- $Sq_{xx} + Sq_{xWidth} - 1 = ImInputWidth$
- All color inside the Sq_x is black

In the chromosome evaluation process, it also records minimum location of Sq_{x_x} and Sq_{x_y} , and maximum location of Sq_{xWidth} and $Sq_{xHeight}$ into variable $Sq_{xMinimX}$, $Sq_{xMinimY}$, $Sq_{xMaxWidth}$, and $Sq_{xMaxHeight}$. This variable will become dynamic parameter in GA process.

3) Chromosome Selection

This process will start from creating chromosome that has smallest objective function. To get this fitness

function we will use is $\frac{1}{(1 + objective\ function)}$. We need to add 1 to avoiding errors caused by divide by 0. After we get all the fitness function for each population, then we will need to find probability for each population

with $\frac{fitness\ function(each\ population)}{total\ fitness\ function}$. From this probability, GA will know which chromosome that has bigger fitness and the function will have bigger chance to be next generation depends on other chromosome.

For chromosome selection, we will use roulette wheel, to do this we need to find cumulative from probability for each population. After we get this result, the next step is to generate random number from 0 until 1 as much as population. We need to compare each random number with each cumulative probability, and the result is the nearest value in from cumulative probability.

4) Cross Over

After finishing with chromosome selection, the next one is to cross over process. One of the methods that are used in this process is one-cut point, that is choosing one of random position in the chromosome parent then exchanging between the gene. The chromosome that is selected by random will become as parent, the number of chromosome in cross over process is affected by crossover rate. This is pseudo-code for crossing over process begin

```

number ← 0;
while(number < populasi) do
  Roulette[number] ← random(0-1);
  if (Roulette[number] < cros_over_rate) then
    select Chromosome[number] as parent;
  end;
  number = number + 1;
end;
end;
```

5) Mutation

Final stage of GA is mutation, the number of chromosome mutations in a population is determined by the mutation rate parameter. The mutation process is performed by substituting one randomly selected gene with a new value obtained randomly. The process is as follows. First, we calculate the total length of gene present in one population. In this case, the total length of the gene is $\text{total_gen} = (\text{number of gene in chromosome}) * \text{population} = 5 * 1 = 5$.

To select the position of the mutated gene is done by generating a random integer number between 1 to total_gen , that is 5. If the random number we generate is smaller than the mutation_rate variable then select that position as a mutated sub-chromosome. In this case, our mutation rate is 80%, then it is expected there are 80% of total_gen experiencing mutation. Number of mutations = $0.8 * 5 = 4$.

From this process, we will get the result from objective function calculation after one generation, the value of the mean result of the objective function is more decreased or increase depends on the which side, the result of objective function before the selection, crossover and mutation. This indicates that the chromosome or solution produced after one generation is better than the previous generation.

These chromosomes will undergo the same process as the previous generation of evaluation, selection, crossover and mutation which will then produce new chromosomes for the next generation using dynamic parameter that already save in variable before. This process will be repeated until a number of generation that has been set before. Figure.3-6 show how the GA with dynamic parameter is executed.

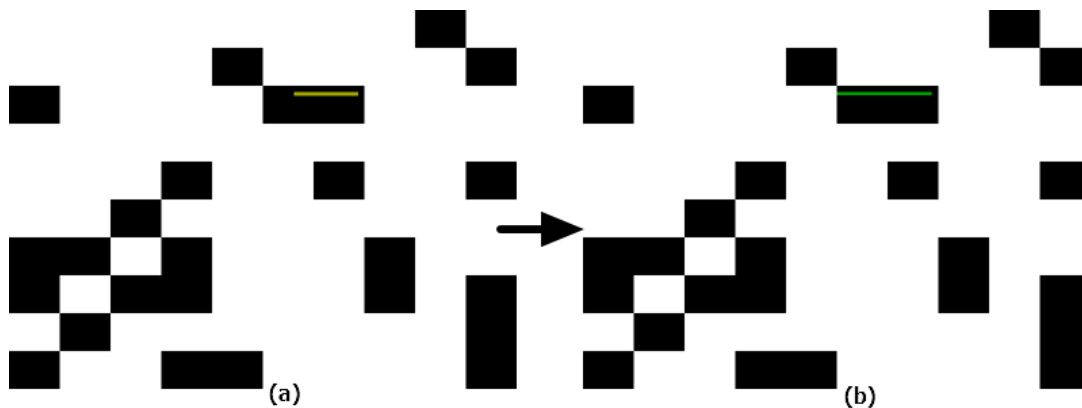


Figure 3. Piece Detection in GA, (a) Start GA (b) result of left edge detection

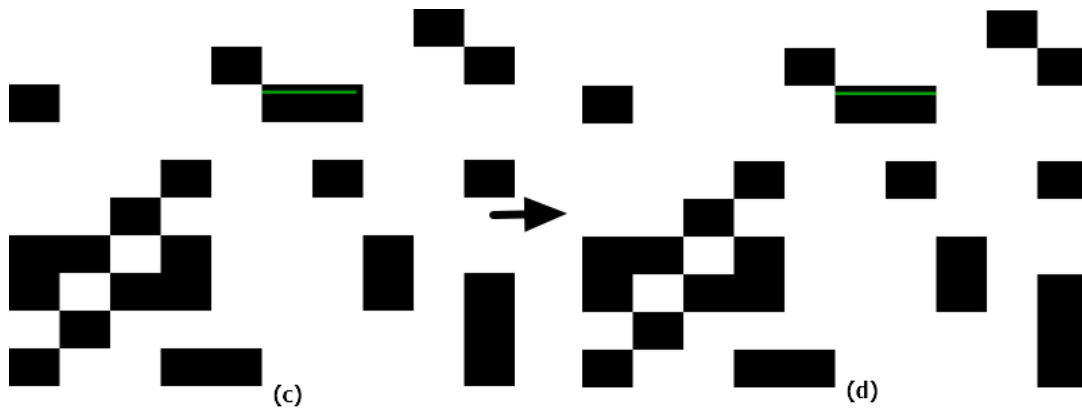


Figure 4. Piece Detection in GA, (c) Start GA (d) result of right edge detection

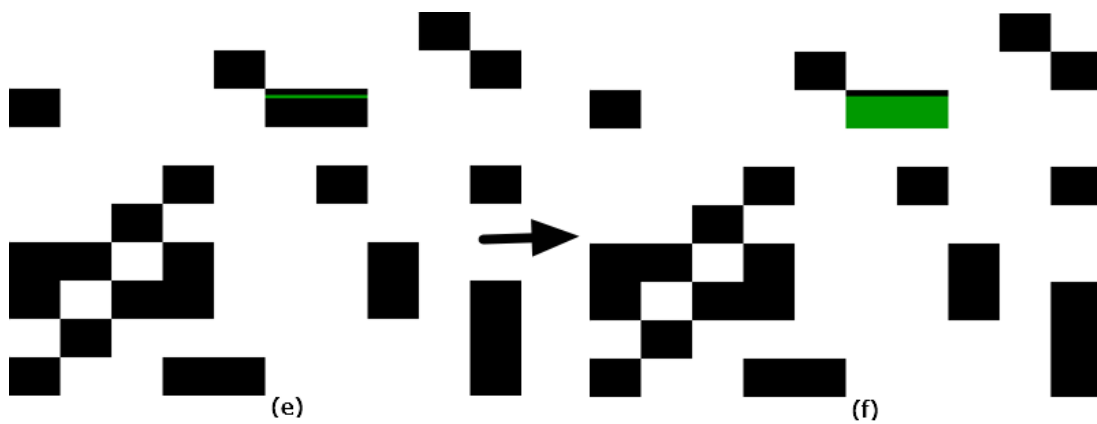


Figure 5. Piece Detection in GA, (e) Start GA (f) result of bottom edge detection

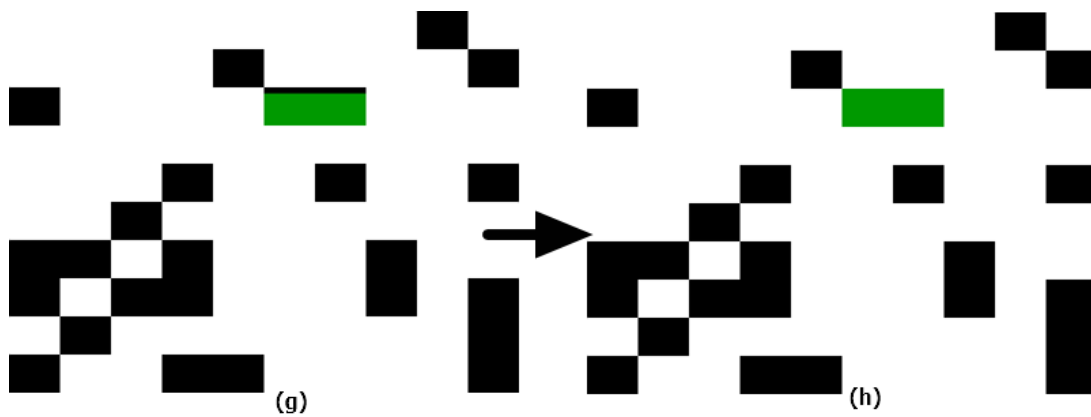


Figure 6. Piece Detection in GA, (g) Start GA (h) result of top edge detection

The final result from Figure 6.(h), all information in green color will be saved into SqN variable.

C. Pieces Mapping

Based on the result from B, we will use SqN as a mold and map it above Im_{BlackWhite} sequentially by making row and column start from the location x,y in (1,1) until width (Sum_{Column}) and height (Sum_{Row}) of Im_{BlackWhite}. The mapping movement starts to fill the row from left to right and column from top to bottom. In the process, it also detects the color inside every SqN that already put above Im_{BlackWhite}, if the color inside the SqN is just black, it will be labeled with “M” for missing piece. If there is just white color, it will be labeled with “P” for NOT missing piece. If they are mixed-color (black and white) inside SqN, system will calculate the area of black and white color. If the total area one of them is more than other, the SqN will be labeled with “E” for error square, see figure.7-8. If in this process, it has label “E”, then we will crop the image from Im_{BlackWhite} with label “E” from this process and makes it to become Im_{input} and repeats the process from point A, for the detail you can see at figure.9-12. Other, if in this process has no label “E”, it means that all processes are already finished, and we will know the location of missing pieces based on the final image (Im_{final}) that mapping with SqN, see figure.13.

$$Sum_{Column} = \left\lceil \frac{Im_{RealWidth}}{Sqx_{Width}} \right\rceil \quad (5)$$

$$Sum_{Row} = \left\lceil \frac{Im_{RealHeight}}{Sqx_{Height}} \right\rceil \quad (6)$$



Figure 7. Piece Mapping

In this mapping, the process we propose three kinds of mapping as follows:

- 1) Based *Piece Mapping with regular size*

This process works by using the size of SqN that is found during the edge detection process. Therefore, the size of SqN_{Width} and SqN_{Height} still same as before and we call this method as GA1.

2) *Piece Mapping with second smallest size*

This process works by taking the remainder of the width of Im_{BlackWhite} with SqN_{Width} and the height of Im_{BlackWhite} with SqN_{Height}. If the result rests smaller and bigger then 0, then SqN_{Width}=mod(Im_{BlackWhite},SqN_{Width}) and SqN_{Height}=mod(Im_{BlackWhite},SqN_{Height}). We call this method as GA2

$$f(\text{mod}(\text{Im}_{\text{BlackWhite}}, \text{SqN}_{\text{Width}})) \begin{cases} = 0 & \text{SqN}_{\text{Width}} = \text{SqN}_{\text{Width}} \\ > 0 & \text{SqN}_{\text{Width}} = \text{mod}(\text{Im}_{\text{BlackWhite}}, \text{SqN}_{\text{Width}}) \end{cases} \quad (7)$$

$$f(\text{mod}(\text{Im}_{\text{BlackWhite}}, \text{SqN}_{\text{Height}})) \begin{cases} = 0 & \text{SqN}_{\text{Height}} = \text{SqN}_{\text{Height}} \\ > 0 & \text{SqN}_{\text{Height}} = \text{mod}(\text{Im}_{\text{BlackWhite}}, \text{SqN}_{\text{Height}}) \end{cases} \quad (8)$$

3) *Piece Mapping with the last smallest size.*

This process works by putting GA2 into repetition until we find the smallest size of square piece. We call this method as GA3.

$$\sum_{\text{SqN}_{\text{Width}} > 0} \text{if}(\text{mod}(\text{Im}_{\text{BlackWhite}}, \text{SqN}_{\text{Width}})) \begin{cases} = 0 & \text{SqN}_{\text{Width}} = \text{SqN}_{\text{Width}} \\ > 0 & \text{SqN}_{\text{Width}} = \text{mod}(\text{Im}_{\text{BlackWhite}}, \text{SqN}_{\text{Width}}) \end{cases} \quad (9)$$

$$\sum_{\text{SqN}_{\text{Height}} > 0} \text{if}(\text{mod}(\text{Im}_{\text{BlackWhite}}, \text{SqN}_{\text{Height}})) \begin{cases} = 0 & \text{SqN}_{\text{Height}} = \text{SqN}_{\text{Height}} \\ > 0 & \text{SqN}_{\text{Height}} = \text{mod}(\text{Im}_{\text{BlackWhite}}, \text{SqN}_{\text{Height}}) \end{cases} \quad (10)$$



Figure 8. Found black color size 50% or more of the mold area

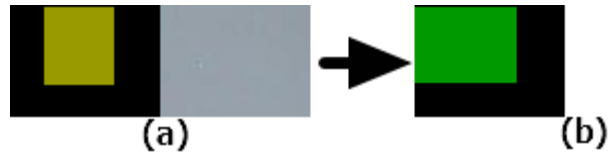


Figure 9. Piece Detection in GA, (a) Start GA (b) result of left edge detection

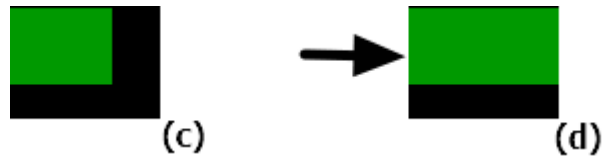


Figure 10. Piece Detection in GA, (c) Start GA (d) result of right edge detection

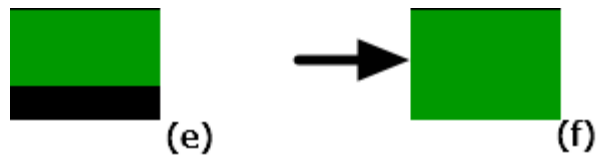


Figure 11. Piece Detection in GA, (e) Start GA (f) result of bottom edge detection



Figure 12. Piece Detection in GA, (g) Start GA (h) result of top edge detection

This process always repeats until we cannot find other missing pieces inside the mold (SqN).

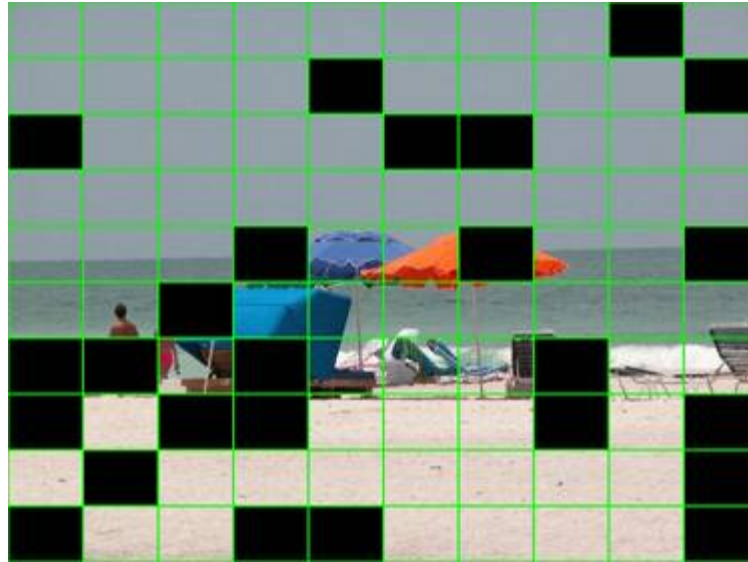


Figure 13. Piece Mapping

D. Labeling Piece

For the labeling piece process, actually this process has already included with the piece mapping in step C, The labeling code is divided into 4

- 1) Label B is used for molding that inside the mold, there is black color but the size is less than 50% of mold area.
- 2) Label M is used for missing pieces / there is just one color that is black.
- 3) Label P is used for pieces / there are other colors than black color.
- 4) Label E is used for mixed-pieces that inside the piece maybe still have pieces and missing pieces that being merged together / there is black color more than 50% of the mold size.

This label will be visible after the system cannot find another missing piece (black color more than 50%) inside the mold SqN. The label and the mold SqN will be merged into Im_{Real} show in figure.14.

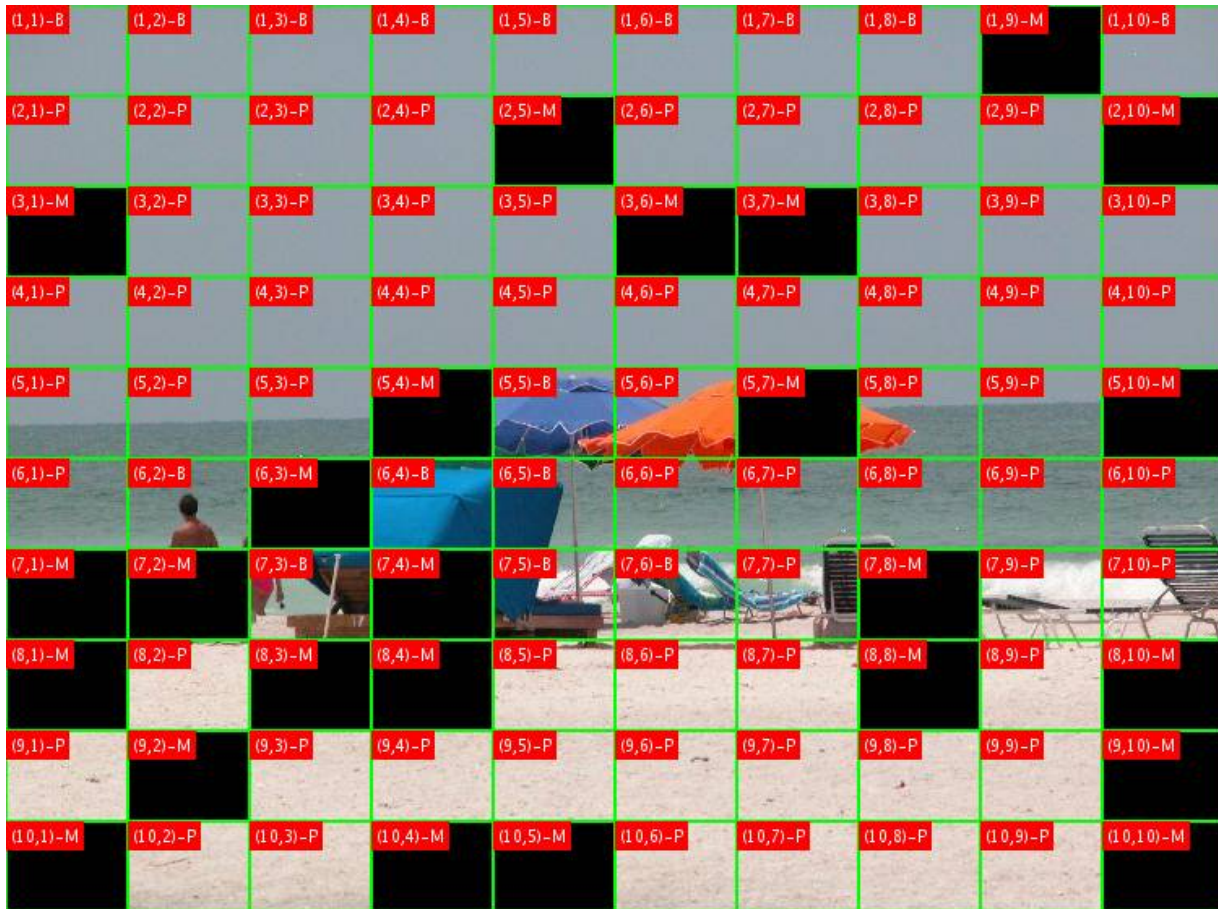


Figure 14. Result of Labeling Piece

III. EVALUATE THE PERFORMANCE OF THE PROPOSED METHOD

Experimental tests have been developed in order to evaluate the performance of the missing pieces detection in the square jigsaw puzzle. Test platform was implemented in MATLAB 2017, on a PC using INTEL® Core™ i5-6300HQ CPU @ 2.30GHz with memory 16GB DDR4 and NVIDIA GeForce GTX960M. These experiments mainly address task such as:

- 1) Random Square with Random Size and Position Generator
- 2) Edge Detection
- 3) Pieces Mapping
- 4) *Labeling*

The image data we use consists of three kind of sizes, that is 756x560 as many as 20 pieces of different images with 100 piece of square jigsaw puzzle in each picture, 980x644 as many as 20 pieces of different images with 1000 piece of square jigsaw puzzle in each picture and 1848x1400 as many as 3 pieces of different images with 10,000 piece of square jigsaw puzzle in each picture. Each size contain 25%, 50%, 75% and 99% of the missing pieces. The result from this experiment we can see at table 2

Table 2. Experiment Result

| Image Size | Square Jigsaw Pieces | Missing Pieces | BALSEM Solving Time | BADPIG | | | | | |
|------------|----------------------|----------------|---------------------|----------------------------|----------|----------|-------------------------|----------|----------|
| | | | | Without Auto Switch System | | | With Auto Switch System | | |
| | | | | GA1 | GA2 | GA3 | GA1 | GA2 | GA3 |
| 756x560 | 100 | 25% | 169.3623 | 20.82991 | 14.29005 | 13.18909 | 18.99259 | 17.16764 | 15.77706 |
| | | 50% | 498.8874 | 23.59254 | 21.35266 | 25.20633 | 18.4549 | 18.60157 | 18.31809 |
| | | 75% | 288.0416 | 26.15561 | 20.92612 | 18.17909 | 21.15221 | 13.20898 | 12.42337 |
| | | 99% | 14.36808 | 55.28177 | 41.5716 | 36.72201 | 15.32156 | 15.63513 | 17.73745 |
| 980x644 | 1000 | 25% | 169.9499 | 22.59884 | 17.44528 | 15.79022 | 22.86243 | 15.20507 | 16.15425 |
| | | 50% | 525.0542 | 25.95535 | 21.53064 | 21.84308 | 17.71973 | 16.74954 | 20.2955 |
| | | 75% | 244.0569 | 31.65011 | 21.91416 | 20.12933 | 20.41992 | 15.78335 | 14.73276 |
| | | 99% | 10.82156 | 63.69629 | 26.54454 | 28.22306 | 15.19213 | 15.44935 | 17.60001 |
| 1848x1400 | 10000 | 25% | 211.2918 | 39.46724 | 34.05926 | 36.32099 | 30.59552 | 31.35177 | 31.65492 |
| | | 50% | 580.6127 | 46.13778 | 30.24438 | 26.66629 | 34.17817 | 33.08472 | 34.7672 |
| | | 75% | 284.5951 | 53.85192 | 36.35977 | 32.73822 | 30.70335 | 29.52085 | 30.35597 |
| | | 99% | 16.884 | 60.56063 | 43.50961 | 47.1762 | 48.01354 | 27.25381 | 32.22724 |

As you can see in the table above, the cell with different color is the fastest time to find and label missing pieces of jigsaw puzzle in every missing pieces. We have already compared from the BALSEM method and BADPIG method with and without ASS. In ASS case, we also divide into GA1, GA2 and GA3. As we can see from the table above, all images with 99% of missing pieces, BALSEM methods are the fastest. BALSEM methods also have auto switch function, when the system detects the black color more than other, BALSEM will be reverted it, that is why the image with 99% of missing pieces can gain faster time to solve it. In the BADPIG method with ASS, we can see that in GA2 part, more pieces in the image, the method can solve faster than image with a few pieces. Otherwise, BADPIG method without ASS, in the GA3 part, they can be used for solving missing pieces faster in the small image size. Figure 15 shows the time it takes to find the missing piece in the square jigsaw puzzle between the methods that are used and the number of missing pieces.

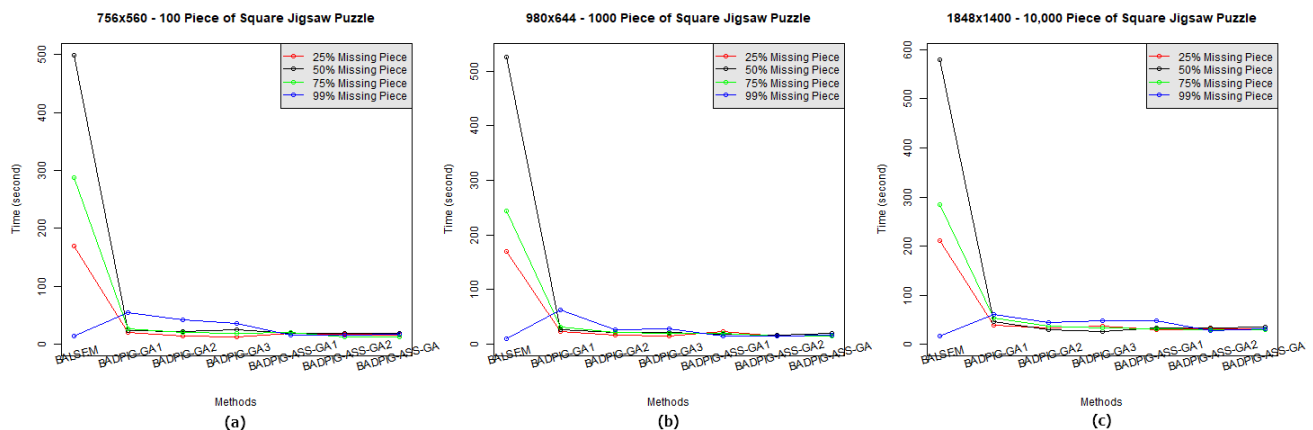


Figure 15. The Timing Of The Resolver Between The Method And The Number Of Missing Pieces

Figure 16 shows the time it takes to find the missing piece in the square jigsaw puzzle between the methods and image size with pieces of jigsaw puzzle.

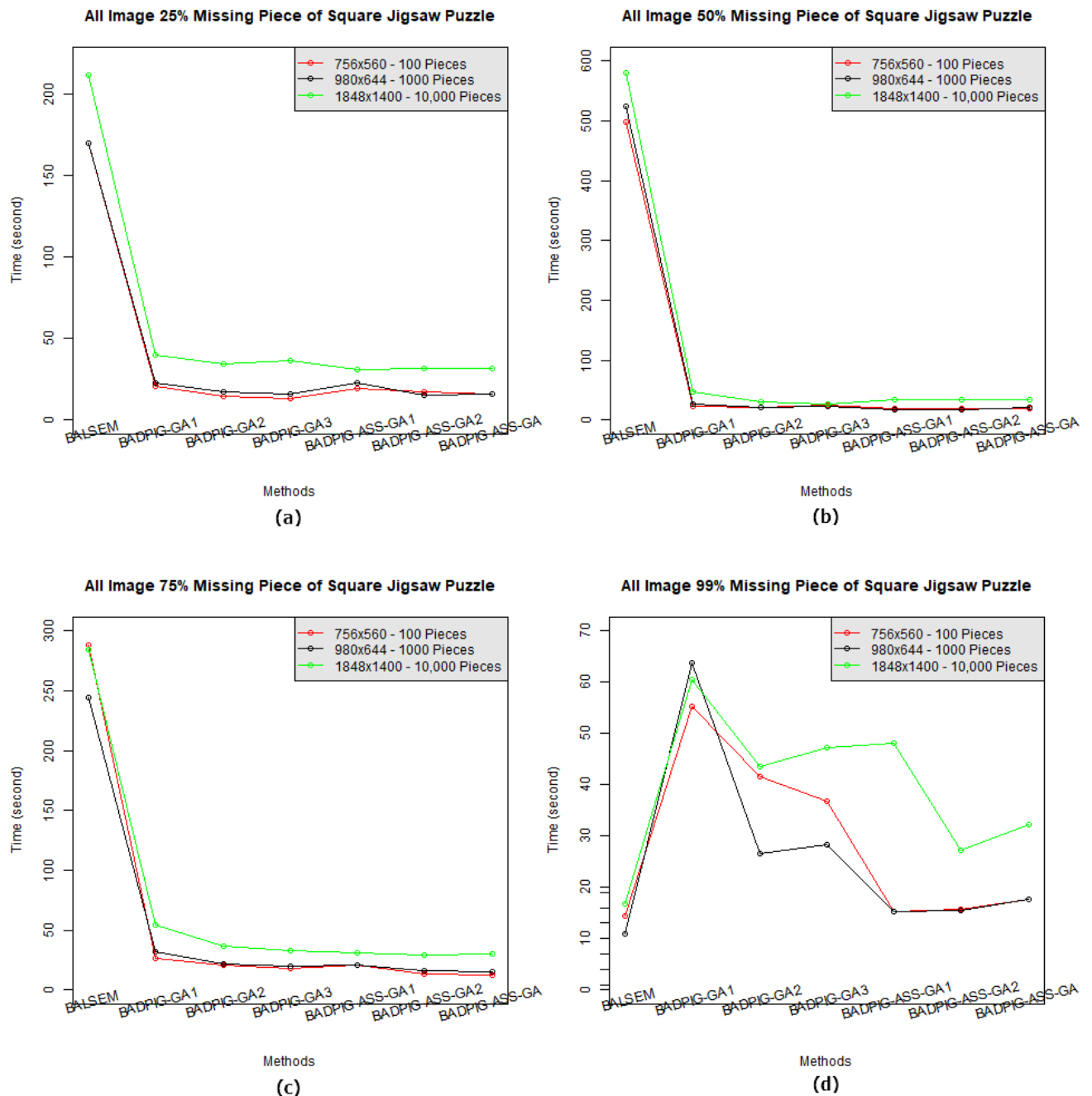


Figure 16. The Timing of The Resolver Between The Method And Image Size With Pieces Of Jigsaw Puzzle

IV. CONCLUSIONS AND PERSPECTIVES ALSO POSSIBLE FUTURE WORKS

This paper shows how to find missing piece in the borderless jigsaw puzzle. The major contribution is a newly proposed BADPIG-ASS, which not only use artificial intelligent that is genetic algorithm and automatic switch system to reduce the time, but also use three kinds method of mapping mold. From this experiment, we draw several conclusions.

- 1) Bigger the image size, more time needed to solve missing pieces with genetic algorithm.
- 2) If the missing pieces are 99%, BALSEM method is much better than genetic algorithm.
- 3) Without ASS in GA1, it consumes more time to solve missing pieces

In the future, we shall further explore other properties of borderless square jigsaw puzzle such as rotation, different shape of pieces other than square. We shall also apply another artificial intelligent to help it solve the problem much faster than this.

REFERENCES

- [1] S. Chen, R. Xia, J. Zhao, Y. Chen, and M. Hu, "A hybrid method for ellipse detection in industrial images," *Pattern Recognit.*, vol. 68, pp. 82–98, 2017.
- [2] Z. Yuan, T. Lu, and C. L. Tan, "Learning discriminated and correlated patches for multi-view object detection using sparse coding," *Pattern Recognit.*, vol. 69, pp. 26–38, 2017.
- [3] O. Starostenko, C. Cruz-Perez, F. Uceda-Ponga, and V. Alarcon-Aquino, "Breaking text-based CAPTCHAs with variable word and character orientation," *Pattern Recognit.*, vol. 48, no. 4, pp. 1097–1108, 2015.
- [4] D. Han and H. S. Hahn, "Axis estimation and grouping of rotationally symmetric object segments," *Pattern Recognit.*, vol. 47, no. 1, pp. 296–312, 2014.
- [5] O. Zaiane, X. Liu, D. Zhao, and W. Li, "A multi-kernel based framework for heterogeneous feature selection and over-sampling for computer-aided detection of pulmonary nodules," *Pattern Recognit.*, vol. 64, no. October 2016, pp. 327–346, 2017.
- [6] Y. Chen, B. Shi, C. D. Smith, and J. Liu, "Nonlinear feature transformation and deep fusion for Alzheimer's disease staging analysis," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9352, no. February 2016, pp. 304–312, 2015.
- [7] Y. Wu, T. Mu, P. Liatsis, and J. Y. Goulermas, "Computation of heterogeneous object co-embeddings from relational measurements," *Pattern Recognit.*, vol. 65, no. December 2016, pp. 146–163, 2017.
- [8] V. Bloom, V. Argyriou, and D. Makris, "Linear latent low dimensional space for online early action recognition and prediction," *Pattern Recognit.*, vol. 72, pp. 532–547, 2017.
- [9] Y. Ji, Y. Chen, H. Fu, and G. Yang, "An EnKF-based scheme to optimize hyper-parameters and features for SVM classifier," *Pattern Recognit.*, vol. 62, pp. 202–213, 2017.
- [10] B. Priego, R. J. Duro, and J. Chanussot, "4DCAF: A temporal approach for denoising hyperspectral image sequences," *Pattern Recognit.*, vol. 72, pp. 433–445, 2017.
- [11] J. Fan, J. Zhang, K. Mei, J. Peng, and L. Gao, "Cost-sensitive learning of hierarchical tree classifiers for large-scale image classification and novel category detection," *Pattern Recognit.*, vol. 48, no. 5, pp. 1673–1687, 2015.
- [12] L. Li, L. Jiao, J. Zhao, R. Shang, and M. Gong, "Quantum-behaved discrete multi-objective particle swarm optimization for complex network clustering," *Pattern Recognit.*, vol. 63, pp. 1–14, 2017.
- [13] M. Swaminathan, P. K. Yadav, O. Piloto, T. Sjöblom, and I. Cheong, "A New Distance Measure for Non-Identical Data with Application to Image Classification," *Pattern Recognit.*, vol. 63, no. October 2016, pp. 384–396, 2016.
- [14] T. Tong, K. Gray, Q. Gao, L. Chen, and D. Rueckert, "Multi-modal classification of Alzheimer's disease using nonlinear graph fusion," *Pattern Recognit.*, vol. 63, no. October 2016, pp. 171–181, 2017.
- [15] C. Zu *et al.*, "Robust multi-atlas label propagation by deep sparse representation," *Pattern Recognit.*, vol. 63, pp. 511–517, 2017.
- [16] F. Albert and N. Aleixos, "Improvements to the TCVD method to segment hand-drawn sketches," *Pattern Recognit.*, vol. 63, no. October 2016, pp. 416–426, 2017.
- [17] X. Zhou, X. Gao, J. Wang, H. Yu, Z. Wang, and Z. Chi, "Eye tracking data guided feature selection for image classification," *Pattern Recognit.*, vol. 63, pp. 56–70, 2017.
- [18] V. Ayala-Ramirez, C. H. Garcia-Capulin, A. Perez-Garcia, and R. E. Sanchez-Yanez, "Circle detection on images using genetic algorithms," *Pattern Recognit. Lett.*, vol. 27, no. 6, pp. 652–657, 2006.
- [19] H. Bostani and M. Sheikhan, "Modification of supervised OPF-based intrusion detection systems using unsupervised learning and social network concept," *Pattern Recognit.*, vol. 62, pp. 56–72, 2017.
- [20] R. Chaieb, K. Kalti, M. M. Luqman, M. Coustaty, J.-M. Ogier, and N. E. Ben Amara, "Fuzzy Generalized Median Graphs Computation: Application to Content-based Document Retrieval," *Pattern Recognit.*, vol. 72, pp. 266–284, 2017.
- [21] L. Tsochatzidis, K. Zagoris, N. Arikidis, A. Karahaliou, L. Costaridou, and I. Pratikakis, "Computer-aided diagnosis of mammographic masses based on a supervised content-based image retrieval approach," *Pattern Recognit.*, vol. 71, pp. 106–117, 2017.
- [22] A. Bahrololoum and H. Nezamabadi-pour, "A multi-expert based framework for automatic image annotation," *Pattern Recognit.*, vol. 61, pp. 169–184, 2017.
- [23] S. Soltanpour, B. Boufama, and Q. M. Jonathan Wu, "A survey of local feature methods for 3D face recognition," *Pattern Recognit.*,

- vol. 72, pp. 391–406, 2017.
- [24] S. Novianto, F. Luo, and K. Penantcha, “Labeling and finding missing pieces of jigsaw puzzle,” in *Proceedings - 2016 International Seminar on Application of Technology for Information and Communication, ISEMANTIC 2016*, 2016.
- [25] S. Novianto and L. Fei, “A Method For Automatic Detection Of The Square Piece In The Borderless Square Jigsaw Puzzle,” *Int. J. Image Process.*, vol. 11, no. 1, pp. 1–11, 2017.

Recognition and Classification of Snack Foods from Cluttered Scene for Dietary Assessment

Amatul Bushra Akhi¹, Farzana Akter² and Tania Khatun³

Department of Computer Science and Engineering

Daffodil International University, Dhaka

Email: akhi.cse@diu.edu.bd, farzana.cse@diu.edu.bd and tania.cse@diu.edu.bd

Abstract—In this paper, we propose an easy approach of identification and classification of high calorie snacks for dietary assessment using machine learning. As an object detection technique we have used point features matching algorithm to identify the object of interest from a cluttered scene. After detecting the object, a Bag of Features (BoF) model is created by extracting Speed up Robust features (SURF) features. This BoF model is used to recognize and classify the snacks items of different categories. We have used three types of snacks images named: Ice-cream, Chips and Chocolate for experimental purpose. Depending on the experimental results our proposed algorithm is able to detect and classify different types of snacks with around 85% accuracy.

Keywords—Point feature matching, BoF, SURF, Food identification.

I. INTRODUCTION

OBESITY is becoming a great health issue day by day. Researchers said that junk foods and processed foods are responsible for increasing the childhood obesity, heart disease, diabetes and other chronic diseases. Brain-derived neurotrophic factor (BDNF) helps in the learning and memory formation of a human brain can be suppressed due to some foods containing high sugar and fat. Eating extra calories can harm the healthy production and functioning of the synapses of our brain. Ice-cream, chips and chocolates are very common and favorite snacks for both child and adults. People often buy these low cost high calorie foods to control their appetite especially when they are busy and unable to take their meal in time. A news was published in "The New England Journal of Medicine" on 2011 reveals that the daily consumption of a single ounce of potato chips led to an average weight gain of 1.69 pounds over four years. On the other hand, a half cup of vanilla ice cream contains around 25mg cholesterol. Excess consumption of high cholesterol can increase our blood-cholesterol levels and increase the risk of heart disease. Chocolate is another source of high fat and sugar which is responsible for acne, obesity, high blood pressure, coronary artery disease, and diabetes. To live a healthy life, it is important to provide more attention in this regards and the great thing is that the public's view towards taking junk food has undergone severe changes day by day. Today's people are more conscious about their health issues and try to maintain a healthy diet.

So an easy but effective calorie measurement techniques can help them to identify the amount of junk food and snacks they

can intake as well as to decide whether the food is harmful or not good for their health. Through our research we try to introduce a new technique for identifying high calorie snacks: Chocolate, Ice cream and chips from our menu and estimate their nutrition value.

II. LITERATURE REVIEW

Many researches have taken place to identify the food and measure the calorie of a food. In our paper we propose methods to recognize some popular snacks such as, chocolate, ice cream and chips. Obesity is conceding a great problem in today's life. The preeminent reason of obesity is consuming more calories than we burn which can seriously undermine the quality of life.

Researchers says, accurately assessing dietary intake is an important factor to reduce this risk. To meet this exigency, Chang et al. (2016) [1] proposed a computer-aided technical method to measure the amount of calorie intake using Convolutional Neural Network (CNN)-based food image recognition algorithm. Probst et al. (2015) [2] is motivated to introduce another prototype for dietary assessment with the help of smart phone as well as the features of image processing and pattern recognition. Scale invariant feature transformation (SIFT), local binary patterns (LBP), color etc. common visual features are used for espying food images. The bag-of-words (BoW) model is used to perceive the images taken by the phone.

Chen et al. (2012) [3] also focused on this major issue and proposed a method of food identification and quantity estimation for dietary assessment. They use Gabor and color features to represent food items. A multi-label SVM classifier combined with multi-class Adaboost algorithm is used to show that the new technique can successfully improve the performance of original SIFT and LBP feature descriptors.

To increase the classification accuracy rate Baxter [4] considered each pixel as a certain ingredient to analyze and classify a food item at the pixel-level, and then using statistics and spatial relationships between those pixels that make up the food as features in an SVM classifier.

Sun et al. (2003) [5] declared a method for investigating the quality of pizza base and tomato sauce spread using computer vision technology. A fuzzy logic system was applied for classifying the sauce spread on the samples.

Hafiz et al. (2016) [6] introduced a method to detect and recognize different types of drinks using vision based algorithms. Thresholding technique is used to segment out and Bag of Feature is used to recognize the drinks.

Kalaivani et al. (2013) [7] proposed a hierarchical grading method applied to the tomatoes to identify of good and bad tomatoes using MATLAB. Thresholding, segmentation and k-means clustering is used to extract feature.

Patel et al. (2011) [8] introduced a multiple features based algorithm to detect the fruit, for efficient feature extraction. The proposed technique can be used for targeting fruits for robotic fruit harvesting.

Recognition of fruit depends on four basic features: intensity, color, shape and texture. In [9], an efficient fusion of color and texture feature is proposed for fruit recognition. Wavelet transformed sub-bands is used to derive some statistical and co-occurrence features for recognition which is done by the minimum distance classifier.

In [10], a vision based techniques have been applied to identify fruits and vegetables diseases. In their research, Rozario et al. (2016) propose a computer vision-based approach for segmentation to identify defected regions from various fruits and vegetables. Savakar et al. (2012) [11] use artificial neural networks for identification and classification of different types of bulk fruit images. Back Propagation Neural Network (BPNN) Algorithm is used to classify and recognize the fruit image samples, using different types of feature set like color, texture, combination of both color and texture features.

III. METHODOLOGY

At the very beginning of our experimental method it is very important to correctly identify the object of interest because our experimental images contain a cluttered scene with different objects. Several preprocessing are required to make the images ready for work properly. Fig 1 shows the complete methodology of our proposed system.

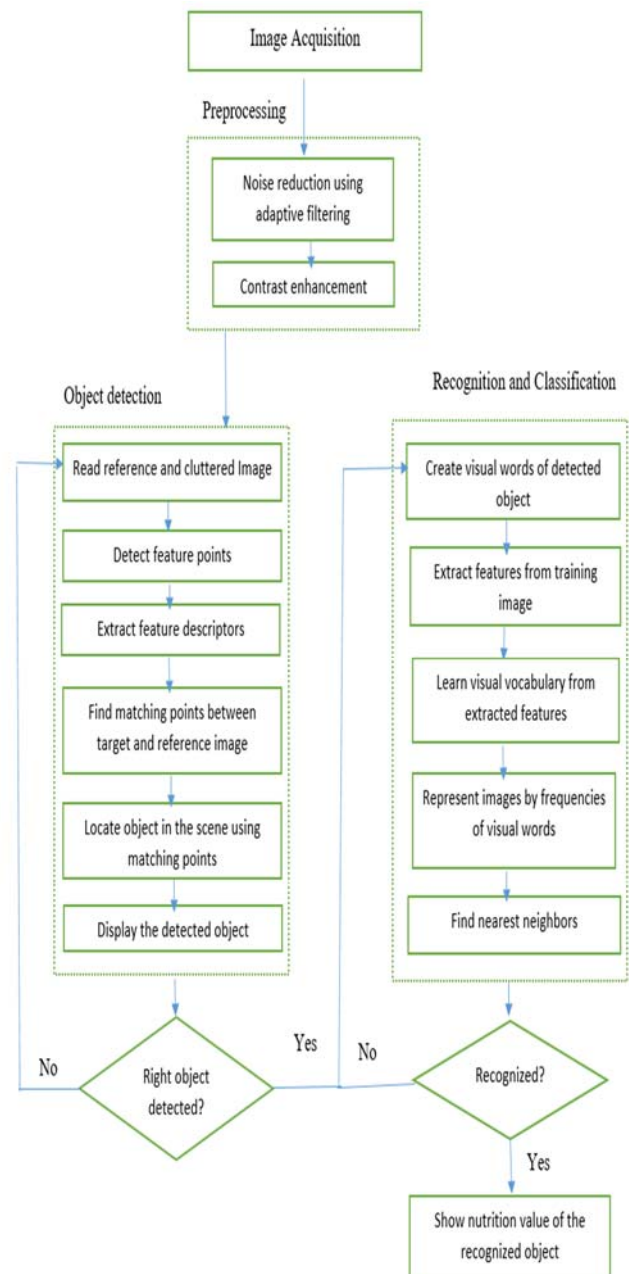


Fig. 1. Block diagram of the proposed methodology

A. Preprocessing

A raw image contains of certain factors such as noise, climatic conditions, poor resolution and unwanted background for which it is not suitable enough to classification and identification. So it is important to improve image quality and prepare the image for further processing to detect the object as accurately as possible. In this paper the pre-processing process consists of noise reduction and contrast enhancement.

1) *Noise reduction using adaptive fuzzy switching median filter:*
The contamination of digital image by salt-and-pepper noise

is largely caused by error in image acquisition. Thus, noise reduction is essential for the accuracy of further processing. In salt-and-pepper noise a certain percentage of individual pixels in digital image are randomly digitized into two extreme intensities. To remove this kind of noise effectively we use two stage noise adaptive fuzzy switching median filter in which salt-and-pepper noise intensities will be identified before identifying the locations of possible noise pixels. Detected “noise pixels” will then be subjected to the second stage of the filtering action, while “noise-free pixels” are retained and left unprocessed.

2) Contrast enhancement:

Image contrast is an important factor which is used to evaluate image quality in addition to distinguish one object from another as well as background. In image processing, contrast enhancement is used to improve the appearance of an image for human visual analysis or subsequent machine analysis. It is created by the difference in luminance reflected from two adjacent surfaces as well as the difference in the color and brightness of the object. In this paper, to contrast the test image we use histogram equalization technique.

B. Detection of Object of Interest

To identify a specific object in a cluttered scene, it is very important to correctly detect the object of interest in the image. Here we have used the most recent and efficient point feature matching algorithm to detect the object we intend to recognize. Here we have used point feature matching algorithm for detecting a specific object based on finding point correspondences between the reference and the target image. This method of object detection works best for objects that exhibit non-repeating texture patterns, which give rise to unique feature matches. This technique is not likely to work well for uniformly-colored objects, or for objects containing repeating patterns

First of all, this algorithm will read the reference image containing the object of interest and a target image containing a cluttered scene and perform feature detection process on both of these images to extract features, which are unique to the objects in the image, so that an object can be detected based on its features in different images. In this algorithm, speeded up robust features (SURF) is used as a patented local feature detector and descriptor. It is partly inspired by the scale-invariant feature transform (SIFT) descriptor but much robust and faster than SIFT. To find points of interest, SURF uses a Hessian matrix based blob detector. The determinant of the Hessian matrix used to evaluate the local change around the points. The points are chosen as feature points where this determinant is maximum. Given a point $X = (x, y)$ in an image I , the Hessian matrix $H(X, \sigma)$ at point X and scale σ , is:

$$H(X, \sigma) = \begin{bmatrix} L_{xx}(X, \sigma) & L_{xy}(X, \sigma) \\ L_{xy}(X, \sigma) & L_{yy}(X, \sigma) \end{bmatrix} \text{-----} (1)$$

Where,

$$L_{xx}(X, \sigma) = I(X) * \frac{\partial^2}{\partial x^2} g(\sigma) \text{-----} (2)$$

$$L_{xy}(X, \sigma) = I(X) * \frac{\partial^2}{\partial xy} g(\sigma) \text{-----} (3)$$

$$\sigma_{approx} = current\ filter\ size \times \left(\frac{base\ filter\ scale}{base\ filter\ size} \right)$$

Where $L_{xx}(X, \sigma)$ in equation (2) is the convolution of the image, $I(x, y)$ at the point x with second derivative of the Gaussian. Non-maximal-suppression of the determinants of the hessian matrices is the main part of the SURF detection process. The main objective of the descriptor is to provide a unique and robust description of features in an image like describing the intensity distribution of the pixels within the surroundings of the point of interest.

After getting the feature descriptors this algorithm finds putative point matches between target image and reference image by comparing the descriptors obtained from both of these images and locate the object with a bounding box in the scene based on the matched putative points. Thus entire object detection process has been completed.

C. Recognition and Classification

This section describes the way to recognize the test image obtained from the cluttered scene and classify the image category by using well established computer vision approach called bag of features (BoF). Object recognition using bag of features is one of the most successful object classification techniques and our target is to classify a given image into one of the pre-determined training objects. The process generates a histogram of visual word occurrences that represent an image which is then used to train an image category classifier.

We have chosen speeded up robust features (SURF) detector to extract features because it is speedup version of SIFT, which uses an approximated DoG and the integral image trick to provide greater scale invariance. If an image is computed into an integral image, using just 6 calculations block subtraction between any 2 blocks can be computed. With this advantages, finding SURF features could be several order faster than the traditional SIFT features.

To train our classifier we have selected a random subset of images from the dataset. To create visual words of detected object k-means clustering algorithm is used on the feature descriptors extracted from training sets. The algorithm iteratively groups the descriptors into k mutually exclusive clusters. The resulting clusters are compact and separated by similar characteristics. Each cluster center represents a feature, or visual word.

The SURF algorithm consists of both feature detection and representation aspects. First, we need to find out the point of interest which we can use for further processing and this step is

called feature detection. Feature detection selects the regions of an image that have unique content, such as corners or blobs. SURF uses a Binary Large Object (BLOB) detector which is based on the Hessian matrix to find the points of interest. Blob is a group of connected pixels in an image that share some common property (e.g.: grayscale value). Feature extraction is done around detected features by transforming a local pixel neighborhood into a compact vector representation. This new representation permits comparison between neighborhoods regardless of changes in scale or orientation.

In order to represent an object, system need to know which part belongs to which. After extracting all the features from all our training images, K-Means clustering is used to group all the features together and extracts the representative centroids. After finding the SURF features of an object we need to find the nearest parts matching these features. So every feature is categorized into one of the part based on the distance from SURF feature to the cluster centers. A histogram is formed for all cluster centers, which represents the frequency of parts for each image. This histogram basically counts the number of times each features has appeared in the given image. We use this to categorize every new image that comes in by computing the Euclidean distance.

IV. EXPERIMENTAL DATA AND RESULTS

Our data set is contrived by three different types of snacks such as chips, chocolate and ice cream collected. To perform this experiment we use 100 chips, 100 chocolate and 100 ice cream images collected from internet. A set of training and testing images are used here where the training set contains 150 images and testing set contains 150 images. We have trained our classifier engine by extracting color, shape, texture and SURF features. Some sample images are given below:



Chocolate Ice-cream Chips

Fig. 2: Sample images from training data set



Chocolate

Ice cream and Chocolate



Chips



Ice cream

Fig. 3: Sample images from validation data set

Our segmentation is performed on different types of images of chips, chocolate and ice-cream after successful identification of desired object. Our proposed system create a classifier depending on the extracted features of identified objects and calculate the nutrition values.

An error matrix also called a confusion matrix is a contingency table that comprise of the information about actual and predicted classifications done by a classification system. Table I to III shows the confusion matrix that appraises the Accuracy rate of the classification. The entries in the matrix are True Positive (TP) rate, True Negative (TN) rate, False Positive (FP) rate, False Negative (FN) rate for each type of dataset. The accuracy (AC) is the ratio of the total number of predictions that were correct. It is derived by the equation:

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \dots \dots \dots (4)$$

TABLE I
CONFUSION MATRIX FOR CHIPS IMAGES

| | | Actual Class | | Accuracy 86 % |
|-----------------|------------------|--------------|-------------|------------------|
| Predicted class | Sample image=300 | Chips | Not Chips | |
| | Chips | 77 (TP) | 24 (FP) | |
| | Not Chips | 17 (FN) | 182 (TN) | |

TABLE II
CONFUSION MATRIX FOR CHOCHOLATE IMAGES

| Predicted class | Actual Class | | | Accuracy |
|-----------------|------------------|-------------|---------------|----------|
| | Sample image=300 | Chocolate s | Not Chocolate | |
| | Chocolate s | 60(TP) | 40(FP) | |
| Not Chocolate | 20 (FN) | 180 (TN) | | 80 % |

TABLE III
CONFUSION MATRIX FOR ICE-CREAM IMAGES

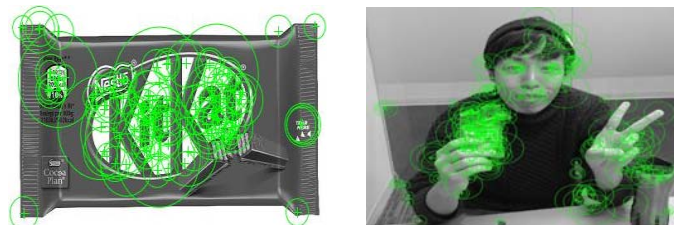
| Predicted class | Actual Class | | | Accuracy |
|-----------------|------------------|-----------|---------------|----------|
| | Sample image=300 | Ice-cream | Not Ice-cream | |
| | Ice-cream | 79(TP) | 22 (FP) | |
| Not Ice-cream | 8 (FN) | 192 (TN) | | 90% |

Fig. 4 represents true detected result of sample image. Our algorithm sometimes provide False Negative and False Positive predicted result. The output image shows this is a chips but the sample image contain a chocolate. This error is occurred due to different texture, shape or color feature. Fig 5 shows step by step output image for detection of object of interest:

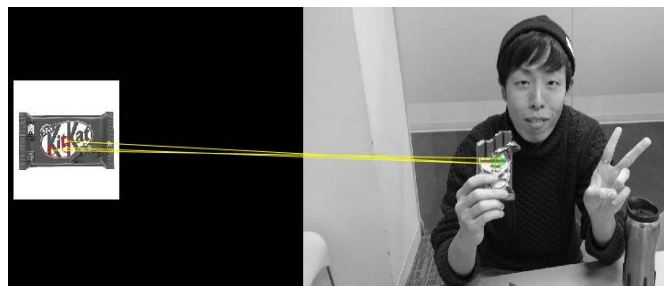


i. Reference Image

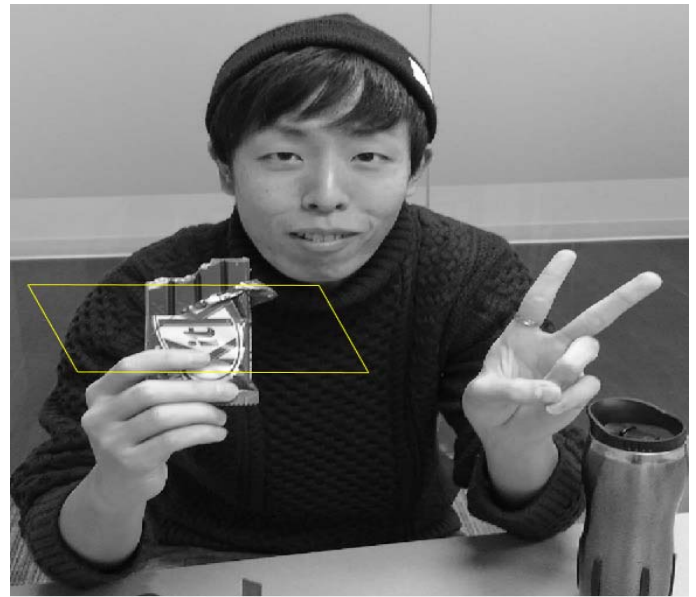
ii. Target Image



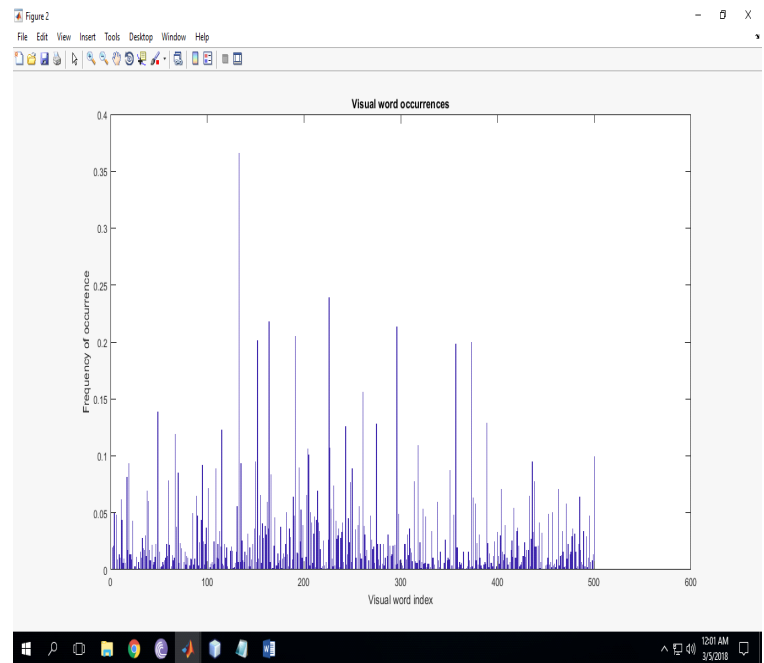
iii. Detect feature points and extract feature descriptors



iv. Matching points between reference and target image



v. Detected Object



vi. Frequency of visual word occurrence

It's a Kit Kat Chocolate

Nutrition value of 100gm Kit Kat Chocolate is

Cholesterol: 11 mg

Fat: 26g

Calories: 518 kcal

Sugar: 49 g

Sodium: 54mg

vii. Final recognition

V. CONCLUSION

In this paper, we proposed a method to classify and to identify high calorie snacks (Ice-cream, Chips and Chocolate) from the test image to measure the amount of calories has taken. Using point feature matching algorithm, our target object is detected in a cluttered scene, with a given reference image of the object. Image category is classified using bag of features approach through finding point correspondences between the reference and the target image. From our experimentation, we found our proposed model is able to detect and classify different types of snacks with around 85% accuracy. The accuracy is good and false positive rate is not so high. People today are very conscious about their health. So, along with the patient, the health conscious person who have a major effect of food calories can be benefitted with this approach. In future, we will try to improve the accuracy by building a robust system which will identify all kinds of snacks more accurately.

REFERENCES

- [1] Liu, C., Cao, Y., Luo, Y., Chen, G., Vokkarane, V., & Ma, Y. (2016, May). Deepfood: Deep learning-based food image recognition for computer-aided dietary assessment. In International Conference on Smart Homes and Health Telematics (pp. 37-48). Springer, Cham.
- [2] Probst, Y., Nguyen, D. T., Tran, M. K., & Li, W. (2015). Dietary assessment on a mobile phone using image processing and pattern recognition techniques: Algorithm design and system prototyping. *Nutrients*, 7(8), 6128-6138.
- [3] Chen, M. Y., Yang, Y. H., Ho, C. J., Wang, S. H., Liu, S. M., Chang, E., ... & Ouhyoung, M. (2012, November). Automatic chinese food identification and quantity estimation. In SIGGRAPH Asia 2012 Technical Briefs (p. 29). ACM.
- [4] Baxter, J. (2012). Food recognition using ingredient-level features.
- [5] Sun, D. W. (2000). Inspecting pizza topping percentage and distribution by a computer vision method. *Journal of food engineering*, 44(4), 245-249.
- [6] Hafiz, R., Islam, S., Khanom, R., & Uddin, M. S. (2016, December). Image based drinks identification for dietary assessment. In Computational Intelligence (IWCI), International Workshop on (pp. 192-197). IEEE.
- [7] Kalaivani, A., Nath, S. S., & Chitrakala, S. (2013). AUTOMATIC dominant region segmentation for natural images. ICCSEA, SPPR, CSIA, WimoA-2013, 273-280.
- [8] Patel, H. N., Jain, R. K., & Joshi, M. V. (2011). Fruit detection using improved multiple features based algorithm. *International journal of computer applications*, 13(2), 1-5.
- [9] Arivazhagan, S., Shebiah, R. N., Nidhyandhan, S. S., & Ganesan, L. (2010). Fruit recognition using color and texture features. *Journal of Emerging Trends in Computing and Information Sciences*, 1(2), 90-94.
- [10] Rozario, L. J., Rahman, T., & Uddin, M. S. (2016). Segmentation of the Region of Defects in Fruits and Vegetables. *International Journal of Computer Science and Information Security*, 14(5), 399.
- [11] Savakar, D. G., & Anami, B. S. (2009). Recognition and classification of food grains, fruits and flowers using machine vision. *International Journal of Food Engineering*, 5(4).



Amatul Bushra Akhi has completed her B.Sc (honors) and M.Sc from the Department of Computer Science and Engineering, Jahangirnagar University. Currently she is working as a lecturer of Computer Science and Engineering Department at Daffodil International University. Her research interest includes image processing, IoT, cyber security, database etc.



Farzana Akter completed her B.Sc (honors) and M.Sc from the Department of Computer Science and Engineering, Jahangirnagar University. Currently she is working as a lecturer of Computer Science and Engineering Department at Daffodil International University. Her research interests includes image processing, Database, Telemedicine, IoT etc.



Tania Khatun has completed her B.Sc (honors) and M.Sc from the Department of Computer Science and Engineering, Jahangirnagar University. She is now working as a lecturer at Daffodil International University. Her research interests includes image processing, machine learning, distributed databases, data mining, telemedicine etc.

Evidence Gathering and Identification of LINE Messenger on Android Device

Imam Riadi

Department of Information Systems
Universitas Ahmad Dahlan
Yogyakarta, Indonesia
imam.riadi@is.uad.ac.id

Abdul Fadlil

Department of Electrical Engineering
Universitas Ahmad Dahlan
Yogyakarta, Indonesia
fadlil@mti.uad.ac.id

Ammar Fauzan

Master of Informatics Engineering
Universitas Ahmad Dahlan
Yogyakarta, Indonesia
fauzan.ammar@gmail.com

Abstract—Smartphone's usage and their applications become popular in our society, nowadays. One of the most influential applications in our social life is the instant messaging application. LINE messenger is one of the popular instant messaging applications around Asian country. LINE has about 60 – 70 percent active users per month from 144 million accounts in Japan, Taiwan, Thailand, and Indonesia. Like most other instant messengers, LINE services are able to keep their user's personal files such as text chats, pictures or photos, and video. These files have the valuables and specific information about the user. In the law enforcement, this kind of information can be an authentic evidence to solve crime cases. In this paper will show the ability of a forensic tool in acquisition digital evidence on Android device. The work is separated into two tests, the application analysis acquisition, and full content acquisition. The digital evidence also has been identified, such as text chats, pictures, the name of the sender and the recipient, and the chat time (timestamp).

Keywords—messenger; evidence; acquisition; forensic; Android

I. INTRODUCTION

Android's smartphone has some interesting applications that popular in our society. One of them is the instant messaging application. It is different from SMS that only provide text message delivery. Instant messaging (IM) applications are able to deliver text messages, pictures, videos, and other files, instantly. There are many names of the instant messaging application based on Android platform. The main factors of its widespread use are because of the ease of use, fun experience, and free cost.

LINE messenger is one of instant messaging application that popular in the Asian country. Exactly 67.3 percent of the monthly active user from 144.7 million accounts in Japan, Taiwan, Thailand, and Indonesia [1]. LINE is basically point-to-point communication system between users. It supports group chat, private chat, and bot chat. Group chat and private chat are for chatting between users while the bot chat is for advertising purpose.

The widespread use of IM application also brings some problems. One of them is cybercrime, especially cyberbullying. Cyber bullying in some social network application is reach

about 25 until 70 percent, while suicide victims around 55 percent [2]. Cybercrime is a serious issue nowadays. Not only bullying, fraud, stalking, and pornographic are also easier occur in IM. It also happens in some instant messaging applications like BBM, Whatsapp, and LINE messenger. According to United Nations's comprehensive study, Cybercrime is a limited number of acts against the confidentiality, integrity, and availability (CIA) of computer data or systems [3]. Figure 1 shows CIA triad that is a guide for measures in information security against cybercrime. It can be said that the information security is the main focus on cybercrime issue.

Cybercrime can be happened in any electronic devices, like Android smartphone. The crime scene in an Android device is able to solve by the investigator with some mobile forensic techniques. Mobile forensics is one of the forensic digital branches that learn on how to perform evidence recovery from a smartphone device [4] Gathering evidences and identify them is one important step to assist law enforcement.

The digital evidence gathered from Android device must be represented as much as possible. The support evidence can be expected to assist law enforcement in solving the cases of digital crimes [5]. The set of information in any Android devices is usually similar. There are Personal Information Management (PIM) applications, messaging, e-mail, and web browsing. NIST [6] mentions 17 potential evidences on the mobile device, such as date/time, text messages, photos,



Figure 1. CIA Triad of information security

outgoing, incoming and missed call logs, instant messaging, etc.

II. LITERATURE REVIEW

A. Digital Evidence

Digital evidence is information stored or transmitted in the binary form that may be relied on in court [7]. Digital evidence is fragile, volatile and vulnerable if it is not handled properly [8]. The change of data can be influenced the result. It is necessary to keep the device in isolation mode. The purpose is to avoid any data from wiping and altering by any condition. The simple move to do this isolation is turned the airplane mode on Android device. Digital evidence can be found in hard drive, flash drive, phones, mobile devices, routers, tablets, and instruments such as GPS [9].

B. Mobile Forensic

Mobile Forensic is a science field that studies the process of digital evidence recovery using the appropriate way from a mobile device. It is the science of recovering digital evidence from a mobile phone under forensically sound conditions using accepted methods [10]. Mobile Forensic is needed because mobile-based (e.g. smartphone device) services are increasing and getting more users, with the growing popularity of mobile computing and mobile commerce [8]. Mobile phone forensic analysis involves either manual or automatic extraction of data to be carried out by the mobile phone forensic examiners [11]. Analyzing digital evidence stored on a Android device is one of mobile forensic challenges in law enforcement.

C. Acquisition and Extraction

Data acquisition from an Android device can be largely divided into the software-based method and hardware-based method [12]. The acquisition is basically a gathering evidence process in order to preserve authentic digital evidence. Extraction is the method to acquire data from the data source. The extraction method can be derived from the physical extraction and logical extraction. Physical extraction is a bit-by-bit copy of the mobile device with the maximum amount of "deleted data or files" recovered [13]. Logical extraction is a method of forensics that principally extracts allocated data from a mobile device and is typically acquired by accessing data in the file system [14].

D. Android

Android is an open-source OS developed by the Google, based on the Linux kernel and designed primarily for touchscreen devices [15]. Android is an operating system created initially for mobile devices, such as smartphones and tablets, but nowadays it has become ubiquitous and popular in other 'smart' devices, e.g., cars, televisions, and watches. Its kernel is Linux-based, but also includes components that are not typically found in a Linux kernel. The Android operating system is a stack of software components which is roughly divided into five sections and four main layers as shown below in the architecture diagram as shown in Figure 2 [16].

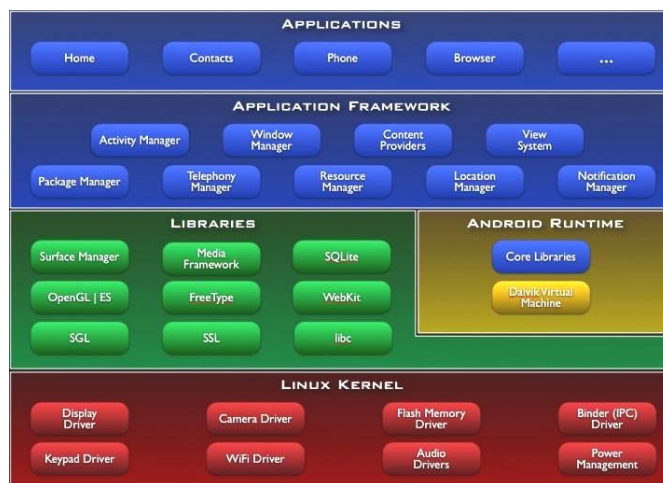


Figure 2. Android architecture

E. LINE messenger

LINE messenger is one of IM provides their users by phone number registration. The users also can create an account in LINE by using a Facebook account. There are many features in LINE messenger, such as private chat, group chat, stickers, and hidden message. Iqbal et al. found that "Hidden messages" feature in LINE are deleted from the device and the LINE servers after the end of the set message timeout duration [17]. So, they thought this feature could be used by criminals to ensure their conversations can be still hidden.

F. MOBILEdit Forensic

MOBILEdit is a forensic tool that allows investigators to logically obtain. This tool uses several connectivity



Figure 3. Screenshots of MOBILEdit Forensic setting : (a)Test 1 (b)Test 2

mechanisms, especially wireless connectivity rather than similar tools. This software is good enough to be used to obtain phone system information and other information such as contacts and text messages. Figure 3 shows about reporting settings in MOBILedit forensic. MOBILedit forensic is one of forensic tool that has been tested by National Institute of Standards Technology (NIST). This tool can run the process of examination, reporting, and logical extraction acquisition [6].

III. TOOLS AND METHODOLOGY

The researchers want to acquire expected digital evidence from LINE messenger on Android device. To ensure the authenticity of the data that has been acquired, recording a hash value on the data imaging results needs to be conducted [18]. In this particular work, the forensic tool and the device are not totally representative of the real condition (cybercrime investigation). The purpose of this work is only to enrich forensic study. Some forensic tool testing might be conducted in CFTT program by NIST [19].

A. Tools

The forensic tool in this research is the main equipment, but it must be supported by other tools to get a good result. The tools that used in this research can be seen in Table 1.

TABLE 1. Tools for forensics research

| No. | Tools | Description |
|-----|--------------------------|--|
| 1. | Workstation | Asus A455L Laptop, Intel Core i3 2.0 GHz Windows OS |
| 2. | Handset / Android Device | Asus Zenfone C Z007 Android ver. 4.4.2, Rooted |
| 3. | USB Cable | USB ver. 2.0 |
| 4. | Forensic Tool | MOBILedit ver. 9.0 |

B. Methodology

The purpose of this research is to gather digital evidence and identify them. The method is using two kinds of extraction techniques. MOBILedit has two kinds of this extraction: application analysis extraction and full content extraction. We want to analyze and identify the two different digital evidences from one forensic tool. Figure 4 show a simulation of data extraction process in the forensic tool.

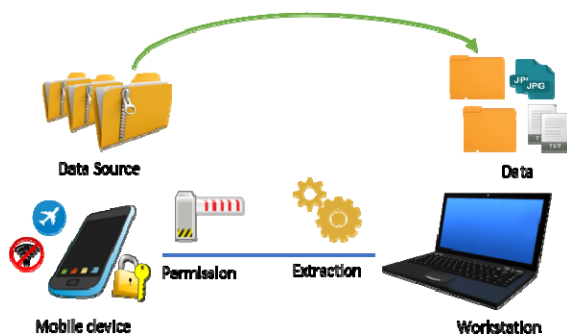


Figure 4. Data extraction process in mobile forensic

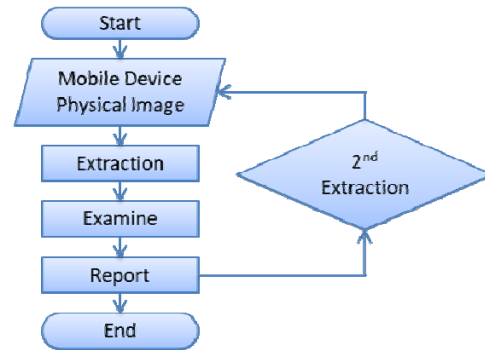


Figure 5. Research methodology

Data extraction process may take some time to complete. First extraction and second extraction processes in Figure 5 show different processing time. It proved by data extraction log. Some forensic tools come with reporting feature, so analyzing and identifying process can be done by observing the report. For a better preparation, prepare working folder on separate media (hard drive) to keep evidence files and data can be recovered or extracted.

IV. RESULT AND DISCUSSION

The result of the process is some potential evidence from two extraction processes: application analysis extraction (extraction I) and full content extraction (extraction II).

IM application that used in this research is LINE messenger version 7.14. As shown in Figure 6, LINE messenger in the mobile device has 125.7 MB data size and 626.7 kB chance size. RAM used in the extraction process on the first test (application extraction) is 53.2 MB. This RAM usage is different from the second test (full content extraction), that is 94.5 MB.

From data extraction log, it is clearly different in the duration of the process. In the first test as seen in Figure 7 (a), data extraction completed in 14 minutes 48 seconds. In the full content extraction, the process complete in 59 minutes 22

| | |
|------------------|-----------------------------|
| | |
| Label | LINE |
| Package | jp.naver.line.android |
| Version | 7.14.0 |
| Application Type | User Application |
| Application Size | 68.3 MB |
| Data Size | 125.7 MB |
| Cache Size | 626.7 kB |
| First Installed | 2017-10-12 05:36:50 (UTC+7) |
| Last Updated | 2017-10-12 05:36:50 (UTC+7) |
| RAM Usage | 53.2 MB |

Figure 6. LINE messenger data extraction in MOBILedit

seconds as seen in Figure 7 (b). Application analysis extraction

Data Extraction Log

```

2017-11-15 06:52:31 Data extraction started - MC
2017-11-15 07:07:19 All 4 archive files were suc
2017-11-15 07:07:19 All 2 audio files were succe
2017-11-15 07:07:19 All 2 documents were succes
2017-11-15 07:07:19 All 14 image files were succ
2017-11-15 07:07:19 All 4 json files were succe
2017-11-15 07:07:19 All 5 sqlite databases were
2017-11-15 07:07:19 All 46 xml files were succe
2017-11-15 07:07:19 All 1434 other files were s
2017-11-15 06:52:34 All 1 applications were succ
2017-11-15 07:07:19 Adb backup was successfully
2017-11-15 07:07:19 Data extraction finished

```

(a)

Data Extraction Log

```

2017-11-15 11:16:32 Data extraction started - MOBILedit Foren
2017-11-15 11:18:21 No phonebook contacts found to extract
2017-11-15 11:18:22 No SIM phonebook contacts found to extrac
2017-11-15 11:18:21 No missed calls found to extract
2017-11-15 11:18:22 No dialed numbers found to extract
2017-11-15 11:18:22 No received calls found to extract
2017-11-15 11:18:21 All 4 messages were successfully extrac
2017-11-15 11:18:21 No organizer events found to extract
2017-11-15 12:15:54 All 143 archive files were successfully e
2017-11-15 12:15:54 All 18 audio files were successfully extr
2017-11-15 12:15:54 All 1 certificates were successfully extr
2017-11-15 12:15:54 Unable to extract all 427 documents - 421
2017-11-15 11:53:22 [read failure] /data/system/dropbox/eve
2017-11-15 11:53:22 [read failure] /data/system/dropbox/eve
2017-11-15 11:53:22 [read failure] /data/system/dropbox/pla
2017-11-15 12:01:16 [read failure] /com.google.android.gms/
tem_cache.db/000067.log
2017-11-15 12:01:16 [read failure] /com.google.android.gms/
can_result_cache.db/000047.log
2017-11-15 12:08:36 [read failure] /com.android.chrome/live
db/000003.log
2017-11-15 12:15:54 All 969 image files were successfully ext
2017-11-15 12:15:54 All 44 json files were successfully extra
2017-11-15 12:15:54 All 545 sqlite databases were successfull
2017-11-15 12:15:54 All 1446 xml files were successfully extr
2017-11-15 12:15:54 Unable to extract all 18215 other files -
2017-11-15 11:32:38 [read failure] /data/data/com.google.an
2017-11-15 11:32:38 [read failure] /data/data/com.google.an
2017-11-15 11:58:42 [read failure] /data/backup/pending/jou
2017-11-15 12:01:28 [read failure] /com.google.android.gms/
2017-11-15 12:01:28 [read failure] /com.google.android.gms/
2017-11-15 12:02:10 [read failure] /com.google.android.gms/
tem_cache.db/MANIFEST-000066
2017-11-15 12:02:10 [read failure] /com.google.android.gms/
can_result_cache.db/MANIFEST-000046
2017-11-15 12:08:38 [read failure] /com.android.chrome/live
db/MANIFEST-000002
2017-11-15 12:15:54 Adb backup was successfully processed
2017-11-15 11:17:47 All 172 applications were successfully ex
2017-11-15 12:15:54 Data extraction finished

```

(b)

Figure 7. Data extraction log from (1) first test and (2) second test.

in MOBILedit only focus on extract any files related to LINE messenger application, such as audio files, documents, image files, SQLite databases, XML files, and other files. In the second test, full content extraction completed its process with more various data, such as phonebook contact, missed calls, incoming and outgoing calls.

In this test, some potential evidence has been acquired totally. The Android device contains LINE messenger artifact such as contact, text messages, picture, audio, and timestamps. Data acquisition on this test uses physical extraction because LINE messenger's data cannot acquire in logical extraction. Rooting on the device meant that the data obtained can be maximally extracted.

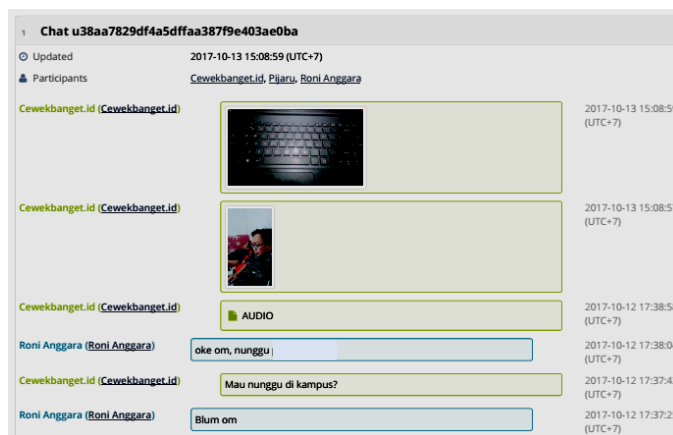


Figure 8. Text message artifact

Potential evidence obtained from both extraction tests shows the difference. Significant differences exist in the number of image and audio files. Both of these proofs may be helpful in cases of the crime requiring a transfer of images or voice mail. While the evidence in the form of text messages obtained fairly complete. However, there is less actuality in reporting the text of the conversation, the sequence or chronology of unordered text conversations as seen in Figure 8. This is according to the researchers is the weakness possessed by MOBILedit as a forensic tool.

The weakness of MOBILedit in chronological order of this text message can be seen in the picture. The disagreement in the chronology of the message text will be a problem in the trial, as it is not actual and weak in constructing arguments. Therefore, the use of other forensic tools needs to be done in order to be a benchmark and present stronger evidence in court. The result of the identification of digital evidence acquired by MOBILedit can be seen in the Table 2.

In addition to the extraction process done in MOBILedit, this tool can perform the reporting process. Reporting done by MOBILedit can be presented in several forms, namely: HTML, PDF, and Excel. While the extraction results can be changed to form Backup file, Export file, and Cellebrite UFDR (for UFED reader). This backup file can be examined repeatedly.

TABLE 2. Evidence Comparison from Data Extraction

| Evidence | Description | |
|---------------------|--------------|----------------------------------|
| | Extraction I | Extraction II |
| Contact Information | 83 contacts | 83 contacts with profile picture |
| Text Message | 51 messages | 51 messages |
| Photos / Images | 14 images | 969 images |
| Audio | 2 files | 18 files |
| Application's File | 1 file | 172 files |

V. CONCLUSION

In this research, the better result produced by full content extraction from MOBILedit forensic tool. Although the text message and timestamp from two reports has similarity, the full content extraction process is able to show more specific data, especially in contact information evidence. Full Content extraction was able to show the profile picture of the contacts. There are various forensic tools that can be used by the examiner to acquire digital evidence. The evaluation of forensic tools can be conducted to get an overview forensic tool ability.

VI. FUTURE WORK

After the researchers know about the ability of MOBILedit forensic to do some extraction processes, the next research about forensic tools must be done. MOBILedit report has some weakness in the data sorting. Maybe other forensic tools have more advantage than MOBILedit. We suggest evaluation of Oxygen forensic or Belkasoft can be conducted in the future work. Both of them are widely used in the mobile forensics.

REFERENCES

- [1] J. T. Quigley, "Despite growing pains, Line made more than \$1b in revenue last year," *techinasia.com*, 2016. [Online]. Available: <https://www.techinasia.com/line-annual-revenue-2015>. [Accessed: 08-May-2018].
- [2] Hariani and I. Riadi, "Detection Of Cyberbullying On Social Media Using Data Mining Techniques," *Int. J. Comput. Sci. Inf. Secur.*, vol. 15, no. 3, pp. 244–250, 2017.
- [3] Conference Support Section, Organized Crime Branch, Division for Treaty Affairs, and Unodc, "Comprehensive Study on Cybercrime," *United Nations Office on Drugs and Crime*, 2013. [Online]. Available: http://www.unodc.org/documents/organized-crime/UNODC_CCPCJ_EG.4_2013/CYBERCRIME_STUDY_210213.pdf. [Accessed: 05-Nov-2017].
- [4] R. Umar, I. Riadi, and G. Maulana, "A Comparative Study of Forensic Tools for WhatsApp Analysis using NIST Measurements," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 12, pp. 69–75, 2017.
- [5] I. Riadi, Sunardi, and A. Firdonsyah, "Forensic Investigation Technique on Android's Blackberry Messenger using NIST Framework," *Int. J. Cyber-Security Digit. Forensics*, vol. 16, no. 4, pp. 198–205, 2017.
- [6] R. Ayers, W. Jansen, and R. Ayers, "Guidelines on Mobile Device Forensics Guidelines on Mobile Device Forensics," *NIST Spec. Publ. 800-101*, 2014.
- [7] D. J. Daniels and S. V. Hart, "Forensic Examination of Digital Evidence: A Guide for Law Enforcement," *U.S. Dep. Justice Off. Justice Programs Natl. Inst. Justice Spec.*, vol. 44, no. 2, pp. 634–111, 2004.
- [8] I. Riadi, R. Umar, and A. Firdonsyah, "Identification Of Digital Evidence On Android's Blackberry Messenger Using NIST Mobile Forensic Method," *Int. J. Comput. Sci. Inf. Secur.*, vol. 3, no. 5, pp. 29–36, 2017.
- [9] M. N. O. Sadiku, M. Tembely, and S. M. Musa, "International Journal of Advanced Research in Digital Forensics," vol. 7, no. 4, pp. 274–276, 2017.
- [10] A. Zareen and S. Baig, "Challenges , Analysis and Tools Classification," no. May, pp. 47–55, 2010.
- [11] K. Curran, A. Robinson, S. Peacocke, and S. Cassidy, "Mobile Phone Forensic Analysis," vol. 2, no. 2, 2010.
- [12] Z. Li, B. Xi, and S. Wu, "Digital forensics and analysis for Android devices," in *International Conference on Computer Science & Education (ICCSE 2016)*, 2016, pp. 496–500.
- [13] J. Kong, "Data Extraction on Mtk-Based Android Mobile Phone Forensics," *J. Digit. Forensics, Secur. Law*, vol. 10, no. 4, pp. 31–42, 2015.
- [14] N. Y. P. Lukito, F. A. Yulianto, and E. Jadied, "Comparison of data acquisition technique using logical extraction method on Unrooted Android Device," *2016 4th Int. Conf. Inf. Commun. Technol. ICoICT 2016*, vol. 4, no. c, 2016.
- [15] M. T. Ahvanooy, P. Q. Li, M. Rabbani, and A. R. Rajput, "A Survey on Smartphones Security: Software Vulnerabilities , Malware , and Attacks," vol. 8, no. 10, pp. 30–45, 2017.
- [16] J. A. Shaheen, M. A. Asghar, and A. Hussain, "Android OS with its Architecture and Android Application with Dalvik Virtual Machine Review," vol. 12, no. 7, pp. 19–30, 2017.
- [17] A. Iqbal, H. Alobaidli, A. Almarzooqi, and A. Jones, "LINE IM app Forensic Analysis," *12th Int. Conf. High-capacity Opt. Networks Enabling/Emerging Technol. (HONET-ICT 2015) poster*, no. IM, 2015.
- [18] A. Prayogo, I. Riadi, and A. Luthfi, "Mobile Forensics Development of Mobile Banking Application using Static Forensic," *Int. J. Comput. Appl.*, vol. 160, no. 1, pp. 5–10, 2017.
- [19] National Institute of Standards and Technology, "Mobile Device Tool Specification Version 2.0," 2016.

An Overview of Biometric Template Security Techniques

Lilian Imali Madivolil

Institute of Computer Science and Information Technology
Jomo Kenyatta University of Agriculture and Technology (JKUAT)
Nairobi, Kenya
madivolil@gmail.com

Abstract—Security remains to be an important aspect in the current age that is characterized with heavy internet usage. People as well as organizations are moving away from ancient authentication schemes like passwords to biometric authentication schemes as they provide better confidentiality and integrity of authentication information. Although biometrics are unique to an individual, they are still susceptible to attacks thus the need to secure them. Several techniques such as encryption, steganography, have been proposed with aim of achieving the security need brought about their use. This paper reviewed some proposed models for biometric template security and their limitation as well as provide recommendation to address some of the limitations outlined.

Keywords—Encryption; Steganography; Biometric; Template Security;

I. INTRODUCTON

A secure authentication system would ensure that only authorized users are able to access or manipulate information. According to Stallings and Brown's study (as cited in [1]) there are four means used in authenticating users; using what they know which involves use of passwords / pin, using what they own which involve tokens, using what they do which involves use of behavioral biometric and using what they possess which are the biometrics.

Although biometric systems are more reliable than other authentication schemes, they are also prone to security threats. As shown in Figure 1, there are eight distinct positions, indicated by letters a – g, where an attack can be launched in a biometric authentication system [2]. Models discussed proposed security mechanisms that target either one or more of the attack points as shown in Figure 1.

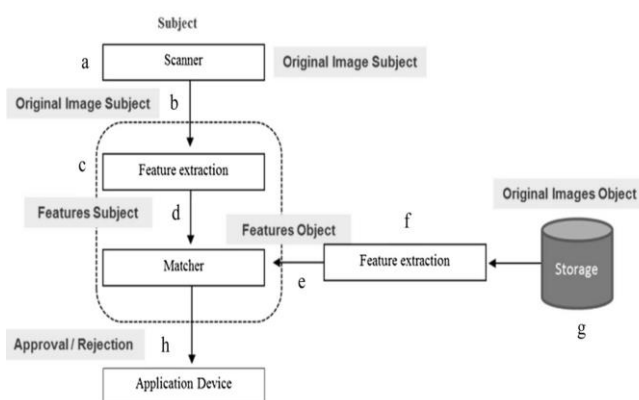


Figure 1 Attack points on a biometric authentication system [2]

The sections following discusses the background knowledge from where the discussed models borrow from (Section 2), then section 3 discusses some of the proposed models and a summary of their limitations and section four concludes with a proposed technique.

II. BACKGROUND OF RELATED WORKS

When securing biometric authentication information, techniques such as encryption, steganography or a combination of these. Other researcher proposed use of multimodal systems as compared to unimodal systems with the previously mentioned techniques so as to heighten security.

A. Cryptography versus information hiding

Cryptography is derived from Greek words kryptos and graphein which means hidden writing. Sensitive information is protected through the use of encryption which transforms the appearance of the actual message without changing its contents. Encryption is an algorithm that changes the plain text to cipher text [3].

Information hiding is an art that involves communication of secret information in an appropriate carrier, e.g., image, video, audio etc. Steganography, an example of information hiding, has at times been confused to cryptography. The difference between these two security techniques is that

steganography conceals the very existence of secret messages being sent whereas in cryptography is used to protect the content of the message by encrypting it [4].

B. Unimodal vs Multimodal biometric systems

Multimodal systems use multiple biometrics characteristics which have to be compared in order to authenticate a user whereas unimodal systems use one biometric trait for the authentication [5]

III. BIOMETRIC TEMPLATE MODELS

A. Biometric cryptosystems

[2] proposed a watermarking technique to secure facial authentication information by combining the use of Principal Component Analysis (PCA) and Discrete Coincise Transform (DCT) which are the most commonly used face regonition and watermark algorithm respectively. In their solution to secure authentication information, they suggest including a timestamp and a logo as watermarks in facial image. During an authentication, a user's facial image is captured for it to be compared to the one that is in database before a grant access or deny access is made by the system. Then the face image is captured, a logo and a timestamp is watermarked onto it. A similar process is performed on the store image in the database. The logo is used as a security measure to distinguish between genuine images and stolen ones whereas the timestamp is used as a session ID. Stolen images that are reintroduced at either points will have a different timestamp from the other one. These two images are then taken through the extraction feature for their respective features to be identified which later on are used in the matching. The proposed solution is used in the scenario where data, in this case face images, are stolen from the system and reintroduced and not in the case where fresh data in reintroduced introduced into the system as well as scenarios where watermarks are included in the data. The approach therefore does not addressed attacks that happen at position a (scanner) and h (approval code). In their experiment, they showed that the face recognition rate is maintained with the embedded watermark as a security mechanism and the watermark detection rate was high without affecting the recognition rate. The research showed that the security of the authentication information was maintained and the PCA and DCT combination did not degrade the performance of the system irrespective of which was used first.

[6] proposed the use of key binding approach to protect biometric template. The basis of their technique is borrowed from randomization technique which is also known as the random masking technique which works by adding adequate large noises to raw values so that individual values cannot be recovered but only statistics of the entire values can be approximately obtained. Yasuda and Sugimura modified this

technique (i.e. randomization method) by taking "lattice points" as noises which is a set of infinite points with sufficient consistency that one can shift any point onto any other point by symmetry of the arrangement. If L represents lattice and you are given a pair (T, K) of a biometric template and a user's specific key, they chose a random lattice point r a subset of L to obtain a "masked" data $H := (T, K) + r$. When a biometric feature Q is queried, it is transformed to a masked data $H' := (Q, 0) + r'$ by independently choosing a random lattice point r' also a subset of L . The difference $H - H' = (T - Q, K) + (r - r')$ also includes the random lattice point $(r - r')$ a subset of L as noise. In this approach, a user and an authentication server are involved. During enrollment *helper data*, H , is generated from T and K (which are the user's template and specific key respectively) using cryptographic tools. These helper data is then stored a secure template in the server's database. When a user is to be authenticated, the correct key K can be extracted from the helper data H only when user's queried biometric template, Q , is close to the original template, T . Then a validity check is performed using the extracted key to output a decision.

[1] proposed use of fully homomorphic encryption (FHE) to secure biometric information during an authentication. In their approach, they propose use of two keys (public and secret keys) to encrypt and decrypt authentication information. Their authentication protocol has three parts; authentication server (AS), client side (CS) and database server (DBS). An assumption they make is that AS and CS are on one site in an organization (Site A) whereas DBS is on Site B. AS is used to generate both sets of keys and manage them. During the enrolment stage, biometric information is obtained, processed then encrypted using a public key to get Bio1 which is stored at the DBS together with the users ID. During the authentication stage, a user presents his/her biometric, which is also encrypted with the public key to obtain Bio1. Bio1 and users ID are then sent to the DBS. Two scenarios exist at the authentication stage. First is the identification phase where Bio1 is taken, but in the DBS, the processes of HD computations and comparisons are done with all records in the DBS. The goal in this scenario is to find which HD result is below the threshold to identify the user ID and grant access or otherwise issue a message reporting that the user is not identified in the current database and is denied access. The second scenario is the verification phase. During this phase, Bio1 and user's ID are taken to the DBS and compared with Bio2 that was obtained during enrolment and stored in DBS. This comparison is made while they are still encrypted. The encrypted result is then taken to AS which decrypts it using the FHE secret key, compares it against the threshold to obtain the final decision to grant/deny access. The matching results (Hamming distance) showed a high-accuracy performance for the iris sample used. It is therefore unlikely that an impostor will be accepted in the biometric authentication system.

Since biometric template is irrevocable if it is stolen, [7] proposed a solution that involves generating cancellable biometrics templates so that the features of the biometric are not revealed, generate a cryptographic key using the cancellable templates from both the sender and the receiver and generate a revocable session key from the biometric traits. Initially, the sender shares two keys, the stego key (K_g) and the shuffle key (K_{shuf}). The stego key (K_g) is generated from a password by sender and receiver using pseudo random number generator (PRNG). The shuffle key (K_{shuf}) is generated randomly which is a binary stream of bits that is stored in a token. Sender shared these two keys with the receiver using asymmetric cryptography. A public key K_{pub} of receiver is used by the sender to encrypt the ($K_{shuf}||pwd$) and sends $EK_{pub}(K_{shuf}||pwd)$ to receiver. Receiver can decrypt the shuffle key and password using his own private key K_{prv} , and they are used for key generation and template sharing, respectively. For session key, they used a biometric-based cryptographic key so as to link users with key. Minutiae points from both the sender's and receiver's fingerprints are extracted then transformed into cancellable templates. Biometrics from both parties involved in the communication are integrated together during the generation of the cryptographic key in order to avoid the use of complex random number generators and eliminate the issue of storing the cryptographic keys. To generate the cryptographic key, both the sender and the receiver share their cancellable template using key-based steganography before being combined together using concatenation-based feature level fusion technique. To randomize how elements are combined, a shuffle key is used. A cryptographic key is then generated using a hash function from the combined biometric. The fingerprints of both parties are not disclosed to either of them. The revocability is provided to the cryptographic key with cancelable template and or with updated shuffle key.

Another proposed scheme is the Multimodal Biometric-based Secured Authentication System using Steganography (MBSASS), which uses two biometrics, say fingerprints and face, to provide message security and user authentication. This system not only protects the message communicated between the users and but also authenticates the sender in an implicit way. From the extracted fingerprint features, the cryptographic key is generated and is shared between the users before the transaction takes place. This key is extracted from fingerprint features that was previously captured and underwent pre-processing to obtain the minutia. This key is obtained using a genetic two-point cross over process. During authentication, the face biometrics of both users are used. In this proposed model, Eigen face-based facial recognition algorithm is used for verification after the facial images undergo pre-processing and shared between the users as well. If user A wants to send the confidential data to user B, the actual message is encrypted by SDES algorithm using the receiver's fingerprint based cryptographic key so as to get the cipher text. The sender's facial image is taken as the cover

image for steganography for embedding the cipher text and the header containing the core point, orientation field value and the number of minutiae points. The generated stego image is divided into several parts depending on the user and then they are scrambled. The order of scrambling is shared among users. The scrambled images and the header are transmitted to the receiver. User B receives the scrambled images in the same order and retrieves the data by first unscrambling the received images and separating the least significant bits from the stego image to get the cipher text and the header. User B then verifies whether the received stego image belongs to the genuine training database by giving that image as the input to the facial recognition algorithm which is transformed into its Eigen face component and for verification it is compared with the mean image. Once authentication is successful, the core point detection algorithm and feature extraction algorithm are applied onto user B's fingerprint image and the related details given in the header and the number of features extracted are found out. If matched, the generated cryptographic key is used to decrypt the cipher text to get the original plain text. By this way, the data security is ensured because B's fingerprint can only decrypt the message. Figure 2.5 shows how this process is executed. Each time user A and B want to share confidential information, different keys are generated and used thus increasing the complexity of the system [8].

In order to secure transactional details such as credit card information of the customers from various attacks such as replay attack, circumvention, repudiation and covert acquisition in e-transactions, a commitment to security is required. [9] proposed a system that enhances E-payment security through Biometric PASS (Personal Authentication using Steganography Scheme) in order to overcome the above attacks. In this system, a B- PASS card is generated by collecting the user's fingerprint and pin number during registration, which is later checked during verification phase. Transaction is possible only if all the three components (fingerprint, pin number and B-PASS card) are available and verified to be genuine. This system gives the internet users the confidence to perform e-transactions without the need to worry about hackers or online shopping frauds.

[10] proposed the use of a 2-factor biometric authentication for a user, i.e. face and voice, and steganography in mobile banking. When the user wants to login into the server or account, s/he is required to enter his eID and password that they were provided with during registration. The server verifies the initial login. If the details provided are correct the user is then redirected to the biometrics authentication page and the user is asked to start video and voice transmission through the mobile phone. During transmission, the video and audio data are hidden into some other images or videos related to normal life. Upon receiving the login information and the stego file, the authentication server decrypts the information to recover the biometric information and attempts to match it

with previously collected biometrics. Upon successful matching and authentication of the user, data transfer happens.

[11] proposed a hybrid to be used in authenticating voters in an online voting process. A voter logs into the system by scanning the face and fingerprint. These biometrics are previously collected and are used in authentication. If authentication is successful, then the voter is allowed to log in to the voting system by entering PIN number and secret key. The system will create the stego image by embedding the secret key and PIN number. The stego image is then sent securely to the server for voter authentication. At the server side, the secret key and PIN number from database stego image and voter's stego image are extracted and compared so as to perform the voter authentication. Once authentication is complete, the voter will be allowed to vote. After casting the vote, the account will be closed and in the database the voted bit will be set for that voter

IV. RESEARCH GAP

Encryption has been used as a mechanism to protect biometric template during transmission in unsecure communication channel. Although this technique offers some level of security, authentication does not happen using the encrypted templates, so they'd have to be decrypted prior to matching. This technique is considered computationally expensive and limits the capacity of large-scale biometric systems from providing a responsive authentication service [12].

Although data in cryptographic techniques are in cypher form, they are plainly visible to the hacker thus stimulate suspicion to the hacker [8]

The use of multimodal biometrics has been used in two aspects; using both in authentication and the other aspect involves concealing the biometric used for the authentication in the other biometric information acquired. In the first scenario, although it provides a significant level of security, the entire process complicates the procedures and protocols as in the case of [13]. Another challenge of using multimodal biometrics for authentication is how to effectively fuse these biometrics effectively [8].

Multimodal biometrics also considered more secure than unimodal biometric systems. The reason for this is enhanced security is that multiple characteristics have to be compared thus making it more difficult for an intruder to trick the system. In such type of authentication systems, a live user has to be present. Though there are benefits, the limitation of multimodal biometric authentication systems is that it presents additional threats to users' data [5]. This security concern is not addressed by some of the systems discussed.

V. CONCLUSION

For the case of [2], emphasis was more on performance while neglecting perceptibility which is one of the reasons why an attacker would steal data if they know it contains useful information. Duration taken to acquire information and authenticate is too long for proposed technique to be used in real life scenarios as in the case of [2] and [1].

In other cases, like that of [6], if the acquired template is not similar or close to be similar to the template stored during the enrollment phase, then a correct key cannot be obtained from the biometric template. A genuine user will be denied access in this case. Other procedures and protocols are complicated such as the case of [13].

To address the challenges outlined above, the use of synthetic fingerprint as a substitute of actual fingerprint to hide biometric data in order to secure the template is recommended. This is because, the time that is need in generating and securing authentication information may be less making the proposed technique suitable in real time. Use of synthetic eliminates need of secure the various biometric information used as well it may reduce the complexity of protocols and procedures used in designing the biometric authentication system. The proposed technique can be coupled with others so as to enhance security of the authentication system.

REFERENCES

- [1] W. A. A. Torres, N. Bhattacharjee and B. Srinivasan, "Privacy-preserving biometrics authentication systems using fully homomorphic encryption," *International Journal of Pervasive Computing and Communications*, pp. 151-168, 2015.
- [2] S. Aljareh, Z. Yusoff and Z. Yusoff, "A watermarking technique to improve the security level in face recognition systems," *Springer*, p. 23805-23833, 2016.
- [3] M. Hussain, A. Wahab and I. Batool, "Secure Password Transmission for Web Applications over Internet using Cryptography and Image Steganography," *International Journal of Security and Its Applications*, pp. 179-188, 2016.
- [4] S. Katzenbeisser and F. A. Peticolas, *Information Hiding Techniques for Steganography and Digital Watermarking*, 2000.
- [5] M. Douglas, K. Bailey, M. Leeney and K. Curran, "An overview of Steganography Techniques Applied to the Protection of Biometric Data," *Multimedia Tools and Applications (Springer)*, pp. 1-41, 13 October 2017.
- [6] M. Yasuda and Y. Sugimura, "Biometric key-binding using lattice masking," *Wiley Online Library*, p. 3405-3414, 2015.
- [7] S. Barman, D. Samanta and S. Chattopadhyay, "Fingerprint-based crypto-biometric system for network security," *EURASIP Journal on Information Security*, pp. 1-17, 2015.
- [8] B. Shanthini and S. Swamynathan, "Journal of Computer Science," *Multimodal Biometric-based Secured Authentication System using Steganography*, pp. 1012-1021, 2012.
- [9] B. Chelliah and S. Geetha, "Enhancing E-Payment Security through Biometric Based Personal Authentication Using Steganography Scheme

– B-PASS," *Recent Trends in Computer Networks and Distributed Systems Security*, pp. 461-472, 2014.

- [10] D. Goyal and S. Wang, "Steganographic Authentications in conjunction with Face and Voice Recognition for Mobile Systems," *ResearchGate*, pp. 1-5, 2016.
- [11] P. Linu and M. N. Anilkumar, " Authentication for Online Voting Using Steganography and Biometrics," *International Journal of Advanced Research in Computer Engineering & Technology*, pp. 26-32, 2012.
- [12] G. Bhatnagar, J. Wu and B. Raman, "Biometric Template Security based on Watermarking," *Elsevier*, pp. 227-235, 2010.
- [13] N. Malkhasyan, "Authentication based on Fingerprint with Steganographic Data Protection," *Internation Journal "Information Theories and Application"*, pp. 289-294, 2013.

AUTHORS PROFILE

Lilian Imali Is a student at JKUAT currently undertaking Master of Science in Computer Systems (2014-2018) and has Bachelor of Science in Mathematics and Computer Science (2012). She has and still is teaching Diploma and Bachelor course at JKUAT and MKU, Kenya. She is also working as an educational consultant at Bada Education Centre, Kenya.

Design and Implementation of A Language Assistant For English – Arabic Texts

A. A. Alabi

Computer Science Department,
College of Technology,
The Polytechnic,
Ile-Ife, Osun, State, Nigeria.
newalabikompurer@gmail.com

O. A. Ayilara

Computer Science Department,
College of Natural Sciences,
Oduduwa University,
Ipetumodu, Ile-Ife, Nigeria
Talk2tobi2004@gmail.com

K. O. Jimoh

Computer Engineering Department
Federal Polytechnic, Ede,
Osun State, Nigeria
oyewumikudirat@gmail.com

ABSTRACT

People across the globe have access to materials such as journals, articles, adverts etc. via the internet. However many of these resources come in diverse nature of languages. Although, English language seems most suitable to most people, some readers do believe that working on materials in one's native language is more enjoyable than in other languages. Researches have shown that Arabic language has not been prominent in terms of online materials and the few existing are most times ignored due to the peculiar nature of its various characters and constructs. Hence, a proper study of its relationship with English language with a view to bringing people closer to its understanding becomes necessary. The system scenarios were modeled and implemented using Unified Modeling Language and Microsoft C# respectively in a way that the expected set of characters of the language of interest was automatically formed with respect to a given input. The procedural steps were properly followed in the development and running of the code using Context-Free Rule Based Technique with the availability of hardware required as clearly described in the design. The system's workability was tested with different source texts as inputs and in each case the resulting outputs were very effective with respect to the translation process. The design here is expected to serve as a tool for assisting beginners in these two languages and so, showcases a one-to-one form of correspondence, hence, more rules and functions for ensuring a more robust are expected in future works.

Keywords: English language, Arabic language, Language Assistant, Machine Translation

1. INTRODUCTION

People across the globe have access to multiple forms of relevant materials such as texts, journals, articles, adverts and websites via the internet. However not all these resources (*especially the text-based*) are in the languages most users understand. Many of these resources come in diverse nature of languages ranging from English language (US and/or British) which is a popular language, French language, Arabic language and a host of others. Many people today believe that reading and working on materials in one's native language seems more enjoyable than other forms of languages thereby calling for the need to translate texts into languages of interest.

Translation as a process, is an activity comprising of the interpretation of the meaning of a text in one language and the production of a new, equivalent text in another; thus, ensuring that both the source and the target texts communicate the same message while taking into account a number constraints such as context, the grammar of the source language, its writing conventions, idioms and the likes. The real challenge in this isn't only about translating the text(s) of a language into equivalent text(s) of another language but the generation of meanings instantaneously and accurately. In real sense, No automated system is meant to replace human translation system but could help in some areas where necessary. For instance, a properly implemented system could assist those seeking to understand certain text materials which aren't written in the language(s) they understand for important decisions making and action taking. Designers in Machine Translation (MT) mostly care about the provision of the meaning of a particular input text in a general form to the user when designing systems. These systems (i.e. Translators) come in different forms; for example, one type functions to provide assistance and help to human translators during translation processes, such as the help in linguistic rules and language grammar. Another type is the automatic system that attempts to directly translate sentences and texts without any human intervention while the third is designed to operate according to different laid down algorithms and translation policies with respect to the contents and contexts of the translation. The key item of interest in this translation mechanism is "languages" which across the globe are

made up constructs derived or created out of syntaxes of different classes. This is simply because of the differences in grammars of the world existing languages. Some languages seem relatively simple and easy to learn to some people while others look difficult to some extent. Meanwhile, information in some languages considered difficult are rarely accessible on net by people; a condition that tends to portray these languages as irrelevant.

Although, some readers and information users have access to many texts, papers, articles and websites over the internet, not all of these texts are in languages of interest, thereby making some people either ignore or neglect some of these materials. Many even find some of these language texts difficult to pronounce and so, making it necessary to get or find the exact translator which could either be a person very familiar with the language or software to do a form of translation into target language before they could understand anything. A good example is Arabic language which most people usually ignore any time they come across while sourcing for information due to its peculiar nature of its various characters and constructs. This, we know could amount to serious waste of time and resources and also loss of credible information in some cases. People do find it extremely uneasy to either source or make reference to Arabic texts due to some reasons and so, making the language less relevant on most of our online platforms. The salient issue here is that, some of the ignored contents could also be of relevance and could probably be the only means of writing and reading especially while dealing with people with only knowledge of Arabic language. The fact still remains that people (i.e. Non-Arabs) have a need to understand Arabic, and Arabs need to understand other languages for handling meaningful materials over the internet, all by an automated system and at no cost. Many non-Arabs today are faced with the need to understand Arabic for the existence of interactive communication in matters of interest. This Arabic language seems very special in terms of its lexis and syntax thus posing serious problems in the area of correspondence for people who do not understand the language. On the other hand, some Arabs (mostly the beginners) sometimes find it difficult to pronounce English text the exact way English speaking people do. What happens in most cases is that the Arabic way (tone) does come into play such whenever people in this regard tend to pronounce letters in English language. Affected people in this regards do improvise by employing a third party for assistance which in most cases tends to turn private information into a public type.

The truth is that no translator can implicitly take care of the syntactic relationships between English and Arabic language because of this peculiar nature of Arabic language but the burden here could drastically be minimized by creating a form of facilitator that could serve as a bridge between the languages in question. The strategy here is to find a form of computer-based language (Rule-Based) tool that could assist in generating any form of English texts into their respective Arabic - based form in a very fast and easy manner of response. This shall increase people with little or no knowledge of the Arabic language in developing interest in the language thus creating a way out for those who want to start working on materials written in such a language. It could also serve as a language assistant for Arabs and others Arabic speakers (people whose sole communication tool is Arabic Language) thus improving their pronunciation of English texts.

2. RELATED WORK

Machine Translation is about the translation of natural languages (Albat & Fritz, 2012); an area which has recently witnessed series of developments. However, lots of works have been pondered in this area of Natural Language Processing; a phenomenon also referred to as Machine Translation (MT). For instance, Nadkarni *et al* (2011) provided a brief description of common machine learning approaches that are been used for diverse Natural Language Processing (NPL) sub-problems. They also discussed how modern NPL architectures are designed with a summary of the Apache Foundation Unstructured Information Management Architecture.

Carbonell *et al* (1981), developed a tool for addressing the several translation problems with examples of English-to-Spanish and English-to-Russian translations. In this, the source was first analyzed and mapped into a language-free conceptual representation with an inference mechanism for the application of contextual world knowledge about items that were only implicit in the input text. The final step then involves the mapping of appropriate sections of the language-free representation into the target language by the natural language generator. Another design was carried out as regards to this but only for the extraction of molecular pathways from journal articles (Friedman *et al*, 2001). Corresponding results from this demonstrated the value of the underlying techniques for the purpose of acquiring valuable knowledge from biological journals. In another development, the use and efficiency of Statistical Machine Translation (SMT) has been lauded then in a model by Shwenk *et al* (2008) comprising of the Open-Source Moses decoder, the integration of a bilingual dictionary and a continuous space target language model for a general purpose French/English statistical machine translation system. Other forms involved the introduction of a Unified Neural Network Architecture and Learning Algorithm by Collobert *et al* (2011). This technique was discovered useful in

various Natural Language Processing tasks including part-of-speech tagging, chunking, named entity recognition and the likes. This system learnt internal representations on the basis of vast amounts of mostly unlabeled training data instead of exploiting man-made input features and was said to be a basis for building a freely available tagging system with good performance and minimal computational requirements. William *et al* (2011), also introduced the Multiscale Geometric Multi-Resolution Analysis (GMRA) for handling the investigation of detection, measurement and modeling techniques to exploit low-dimensional intrinsic structures with a view to improve procedure such as machine learning. The results obtained showed that the approximation error of the GMRA is completely independent of the ambient dimension; thus establishing GMRA as a provably fast algorithm for dictionary learning with approximation and guarantees. Folajimi and Isaac (2012) came up a tool for understanding Yoruba Language using Statistical Machine Translation (SMT). The software employs a machine translation paradigm where translations are generated on the basis of statistical models whose parameters are derived from the analysis of bilingual text corpora. It was discovered that SMT seems to be a veritable tool for translating between English Language and Yoruba Language because of the non-existence of parallel corpus between the two languages. In their work (Kaufman *et al*, 2016), a technique known as Generic Notions of Complexity for two dominant frameworks was introduced as an improvement on the stochastic multi-armed bandit model which proved significantly positive. Also, Maggioni (2016), extended work on Multiscale Geometric Multi-Resolution Analysis (GMRA) application-wise using a Non-Asymptotic Bounds and Robustness procedure (Maggioni, 2016).

However, several other research works were carried out and the trend still evolves on a daily basis. This is as a result of the different forms of scope embedded in most available publications. While certain people argue on the need for totally new designs, some prefer improving on the techniques demonstrated in existing publications. An example of this was the study to determine the relationship between grammar efficacy and grammar performance among Arabic learners on aspects such as Correction of grammar errors, Vocalization of words, and Construction of sentences through questionnaire and it was observed that a moderate exist correlation between grammar efficacy and grammar performance with efficacy of sentence construction as the most noticeable result (Mustapha, 2017).

3. SYSTEM DESIGN

3.1 Methodology

The work here work is designed to replace the involvement of humans in the translation of English (source language) to Arabic (target language). In most cases, the system needs to emulate the thinking strategy as humans do during translation. The various characters sets making up the alphabets in both English and Arabic languages (fig. 1) are to be identified and analyzed. A database of the English alphabets is to be created and made to represent the set of characters representing the source with those of Arabic language taken in to consideration. Also, the syntaxes of both languages are to be made reference to with the aid of in-built tools and some other forms of program segments when needed in terms of their relationship character-wise. The various scenarios in the design of the system were modeled using Unified Modeling Language (UML). For instance, the modeling section involving the Use-Case diagram (figure 2) has to with the demonstrations of the relationship among the various entities making up the system in terms of functions and possibly dependencies. A logical way of representing this user-data relationship is also shown in the class diagram (figure 3). Another aspect of this modeling includes the procedural flow among the various class objects involved in the translation mechanism. The said scenario is diagrammatically illustrated using the Activity diagram (figure 4).

The program is designed in such a way that the expected set of characters (word, sentence etc.) of the language of interest (target language) is automatically formed with respect to a call (an instruction code) by the user with English language as the source language and Arabic language as the target language. This was made possible with the aid of Microsoft C#. So many programming languages were considered in the cause of designing the system. A lot of factors were put into consideration which includes database access, data transmission via networks, database security, database retrieval, multi user network access, data capture, etc. The choice “Microsoft C# (C sharp)” programming language was made to achieve the above set of objectives. Microsoft C# programming language is a user friendly platform that gives room for the design of an interface that can be modified programmatically. The language has the advantage of easy development, flexibility, and it has the ability of providing the developer/programmer with possible hints and it produces a beautiful graphical interface.

| | |
|--------------------------|-------------|
| a | ا |
| b | ب |
| c | س |
| c in front of consonants | ط |
| d | د - د |
| e | ه - ا - ي - |
| f | ف |
| g | غ |
| h | ح |
| ch, sh | ك |
| th | ت |
| i | ي |
| j | ج |
| k | ك |
| l | ل |
| m | م |
| n | ن |
| o | و |
| p | پ |
| q | ظ |
| r | ر |
| s | س |
| s between two vowels | ز |
| th | ث |
| t | ت |
| v | ف |
| w | وا - و |
| x | ا - كس |
| y | واي |
| z | ز |

Figure 1: English - Arabic Alphabets

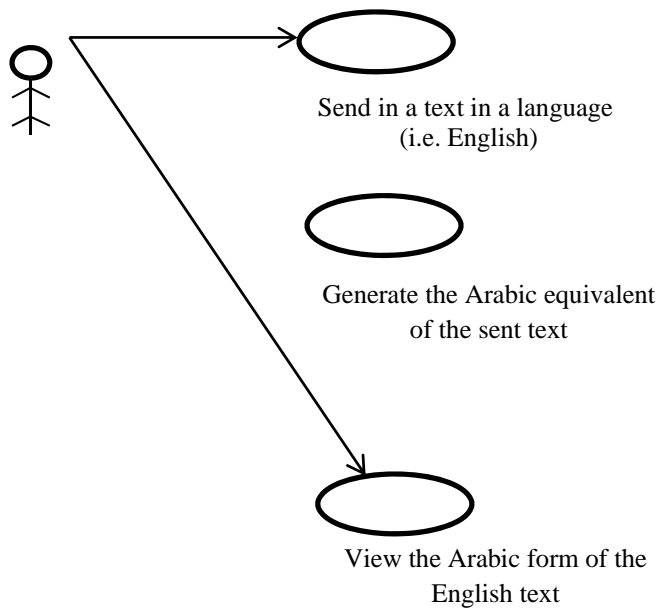


Figure 2: Use-Case Diagram

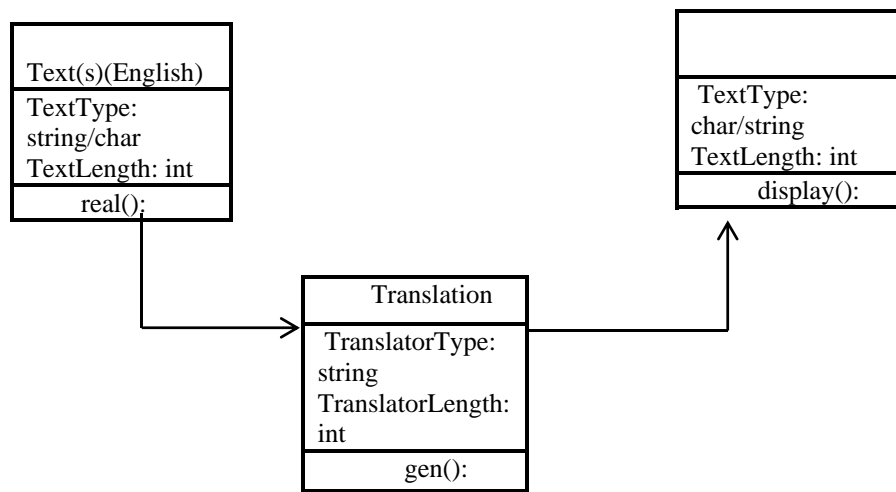


Figure 3: Class Diagram

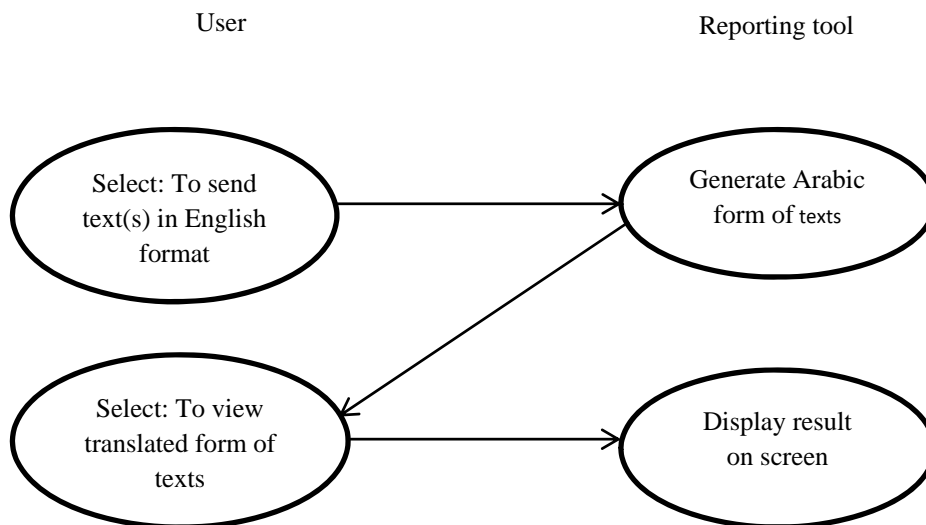


Figure 4: Activity diagram

3.2 Input Design and Output Specification

It is necessary to denote that data inputted in the computer for processing determines what the output usually is. Screen designs are generally or basically made for data entry or capture. With English text as input, the output section of the system in this research work is designed to automatically display its result in Arabic language (*based on the input supplied*); i.e. generate response immediately after the input is received by the system. However, the nature of the output strictly depends on that of the input as well as the correctness of the code implementation with respect to the rules governing the writing and transformation of both character and words in the concerned languages.

3.3 Program Design and Specification

The system's structure, as shown in the flowchart (*fig. 5*) has the general form as a task divided into several sub-tasks, which come together to give the solution to the problem with the translation process as the core stage (*fig. 6*). The program is designed with the specification of having two languages modules namely English Language and Arabic Language.

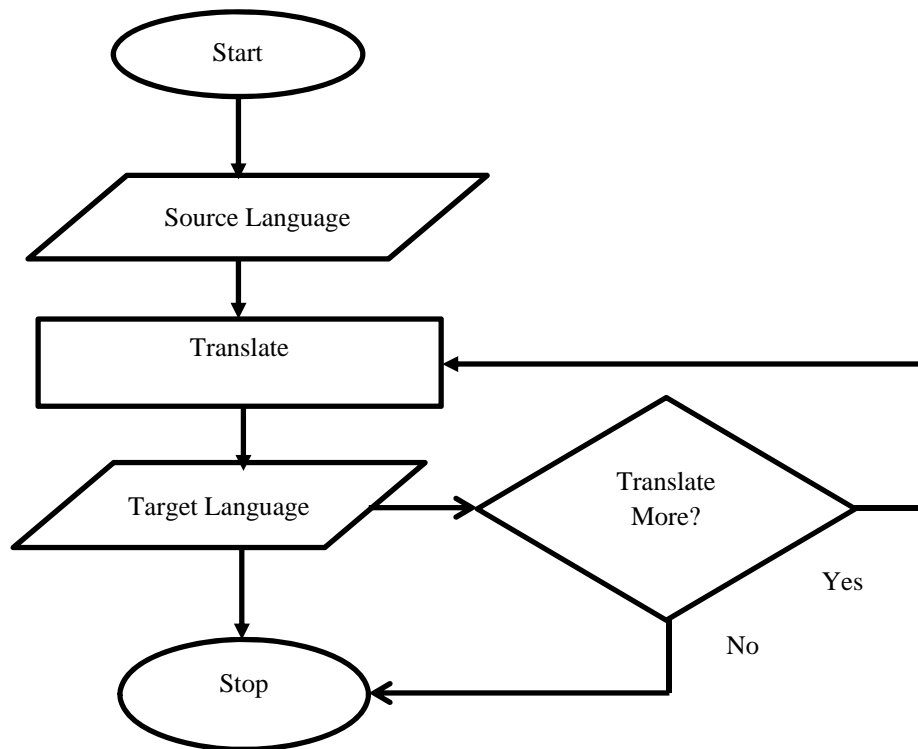


Figure 5: System Flowchart

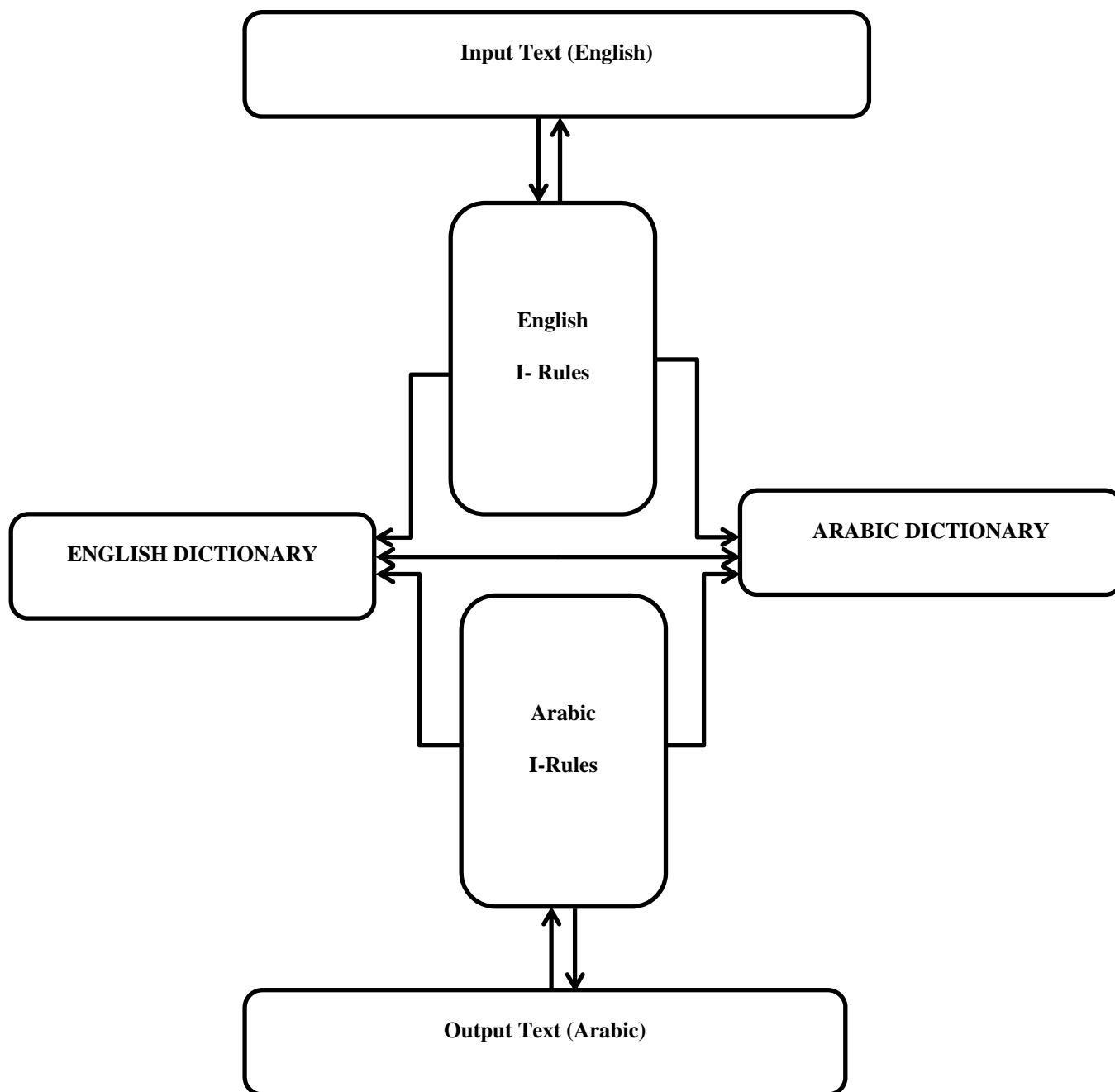


Figure 6: Translation Process

4. SYSTEM IMPLEMENTATION AND EVALUATION

The procedural steps were properly followed in the development and running of the code with the availability of hardware required as clearly described in the design section. The resulting outputs were very effective with respect to the translation process. The system's workability was tested with three different source texts as inputs and in each case, the result came out accurately. For example, the first three sets of English characters (alphabets a, b and c) where the first point of call for translation .This is captured in figure 7; with its first part (upper part) showing the inputs (in English) while the second part (lower part) displays results (in Arabic). Sufficed to that was the translation of some phrases as shown in figures 8 and 9. The various displays in the translation exercise also came up with the processing speed in each of the instances.

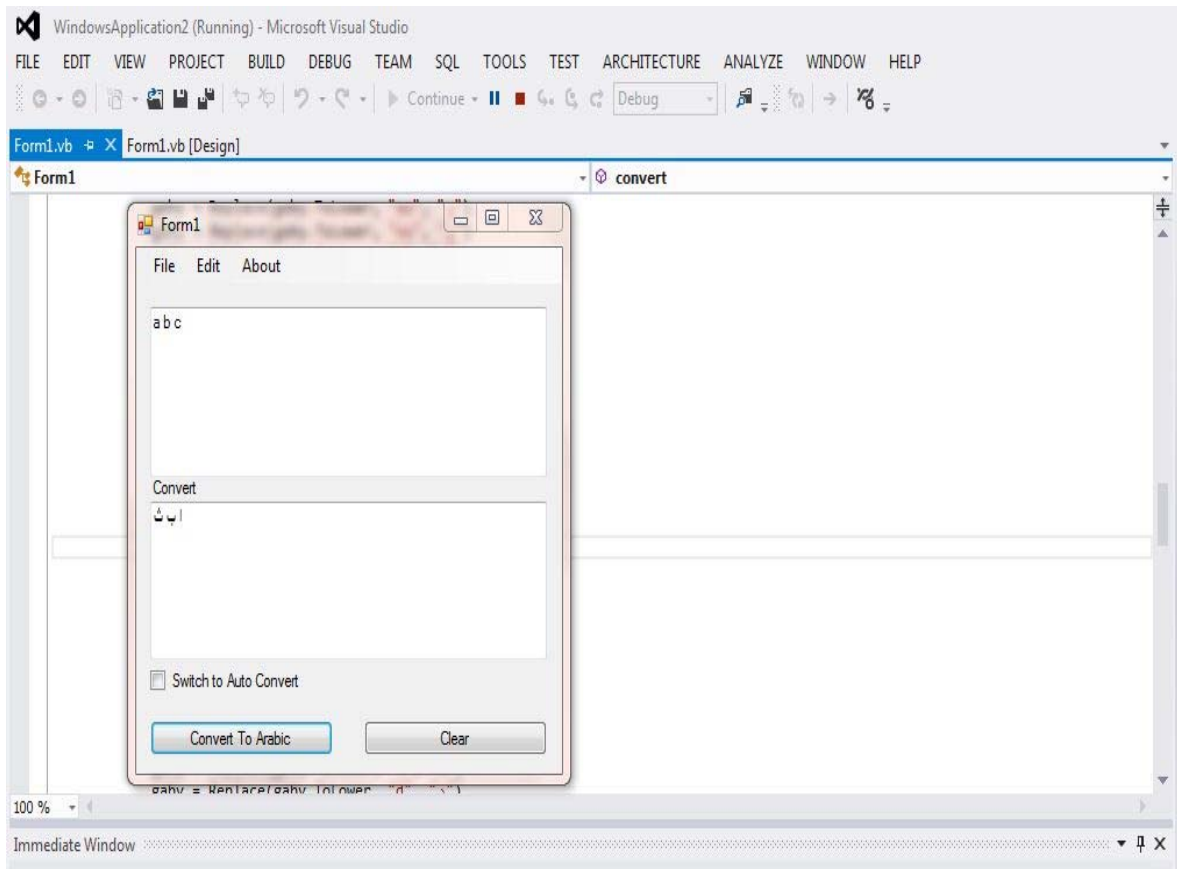


Figure 7: Character Translation

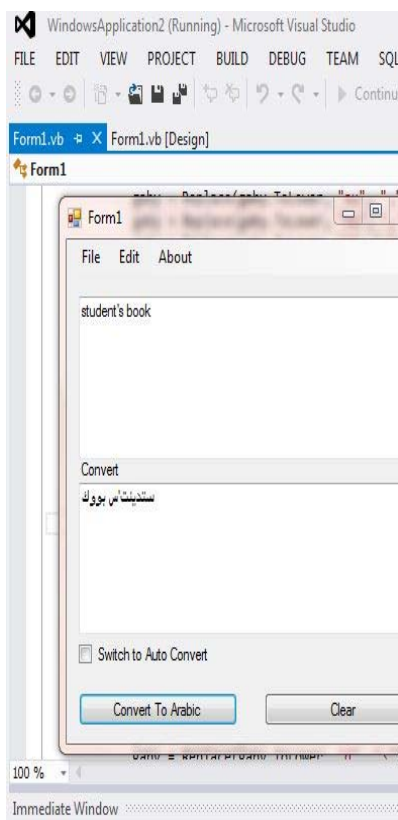


Figure 8: Translation: phrase 1

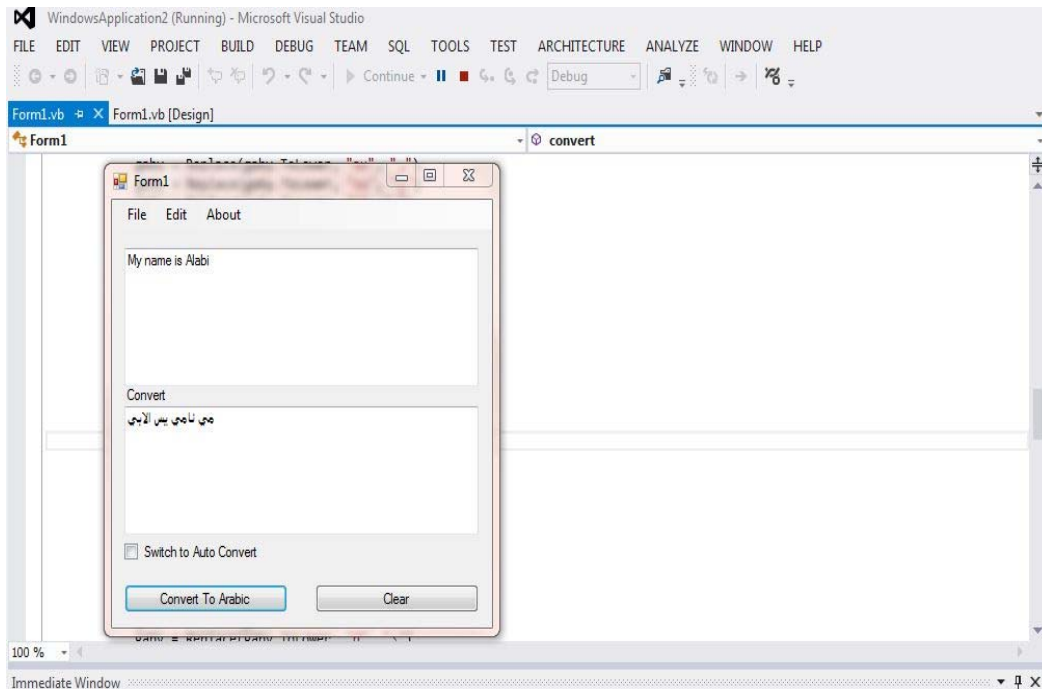


Figure 9: Translation Phrase 2

5. FURTHER ISSUES AND

CONCLUSION

The parameters of interest in this study as demonstrated in the various results obtained show that the system was simply meant to handle the text-to-text conversion between English language and Arabic language. This showcases the form of representation one expects as regards to the Arabic form of a given English text. However, some improvements are clearly necessary to increase both the quality of the study scenario with a view to addressing the possible meaning of the Arabic form of any given English text and possibly make it speech based. This is simply because of the draconian nature of spellings and pronunciations involved in Arabic language. To achieve this, a robust system that takes care of both texts as well as the voice part is required. This shall take care of the targeted goal in future designs as regards to this area of Natural Language Processing.

The developed system proved very efficient with respect to its outputs and the translation was very accurate within reasonable time frame. This represents a useful tool for assisting those seeking redress in a form of assistance in the lexical understanding of texts between English language and Arabic language. Understanding how characters in English are written in their Arabic and vice-versa shall promote a form of familiarity between the two languages thus assisting stakeholders in not just writing of letters but knowing the exact of any character thus enhancing how certain letters are to be pronounced in their respective languages. The different tasks accomplished in the process of designing and implementing the system were properly displayed in the various figures shown and their needed explanations included accordingly. The design here showcases a translation form considered as “One-to-One Correspondence” and so, some future works are still required as stated earlier for the incorporation of more rules as well as functions for ensuring a more robust and more useful form of language translation.

In theory, one is expected to get increased in comprehending a particular language, therefore, getting use to the Arabic form of English letters is expected to create a form of familiarity that could assist Arabs who sometimes find it uneasy to either write or pronounce such texts. This could also be a tool for beginners in Arabic language.

6. REFERENCES

- [1] Albat A. and Fritz T., (2012) . Systems and Methods for Automatically Estimating a Translation Tie. US Patent 0185235.
- [2] Carbonell J. G., Cullingford R.E. and Gershman A. V., (1981). Steps Towards Knowledge-Based Machine Translation. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. PAM 1-3, No. 4.
- [3] Collobert R., Weston J., Bottou L., Karlen M., Kavukcuoglu K. and Kulkarni P., (2011). Natural Language Processing (Almost) From Scratch. Journal of Machine Learning Research. 12, 2493-2537
- [4] Folajimi Y. O. and Isaac O., (2012). Using Statistical Machine Translation (SMT) as a Language Translation Tool for Understanding Yoruba Language. EIE Second International Conference; Computer, Energy Net., Robotics and Telecom. eleCon2012.
- [5] Friedman C, Kra P., Yu H., Krauthammer M. and Rzhetsky A., (2001). GENIES: A Natural-Language Processing System for the Extraction of Molecular Pathways from Journal Articles. BIOINFORMATICS. Oxford University Press. Vol. 17, Suppl.1.
- [6] Kaufman E., Cappe O. and Garivier A., (2016). On The Complexity of Best-Arm Identification in Multi-Armed Bandit Models. Journal of Machine Learning Research. 17, 1-42.
- [7] Maggioni M., Minsker S. and Strawn N., (2016). Multiscale Dictionary Learning: Non-Asymptotic Bounds and Robustness. Journal of Machine Learning Research. 17.
- [8] Mustapha N. F. (2017). Grammar Efficacy and Grammar Performance: An Exploratory Study on Arabic Learners. Mediterranean Journal of Social Sciences, Vol 8 No 4.
- [9] Nadkarni P., Ohno-Machado L and Chapman W. (2011). Natural Language Processing: An Introduction

Journal of The American Medical Informatics Association.; 18:544-551. doi: 101136.

- [10] Shwenk H. Fouet J. and Senellart J., (2008). First Step towards a general-purpose French-English Statistical Machine Translation System Proceedings of the third Workshop on Statistical Machine Translation pages 119-122 Columbus.
- [11] William K., Chen G. and Maggioni M., (2011). Multiscale Geometric Method for Data Sets II: Geometric Analysis. Cornell University Library, New York, USA

Mammographic Image Enhancement using Digital Image Processing Technique

Ardymulya Iswardani
Information System Department
Stmik Duta Bangsa
Surakarta, Indonesia
ardymulya@stmikdb.ac.id

Wahyu Hidayat
Informatics Department
Stmik Duta Bangsa
SURAKARTA, Indonesia
wahyu_hidayat@stmikdb.ac.id

Abstract— **PURPOSES:** this study aims to perform microcalcification detection by performing image enhancement in mammography image by using transformation of negative image and histogram equalization. **METHOD:** image mammography with .pgm format changed to .jpg format then processed into negative image result then processed again using histogram equalization. **RESULT:** the results of the image enhancement process using negative image techniques and equalization histograms are compared and validated with MSE and PSNR on each mammographic image. **CONCLUSION:** Image enhancement process on mammography image can be done, however there are only some image that have improved quality, this affected by threshold usage, which have important role to get better visualization on mammographic image.

Keywords-component; Image enhancement, image negative, histogram equalization, mammographic, breast cancer

I. INTRODUCTION

This section described the motivation background of the studies as follows:

Difficulties in early identification of cancer cell existencies affected by it natural ability to multiply, survive, spread and hide for a certain time [1]. Mamographic screening is the best method for early identification. This method use X-ray to check the patient organs [2], [3]. Cancer can be identified from the presence of microcalcification, microcalcification is a major feature of cancer, however, false identification and unable to get important clues of microcalcification presence often occur [1]–[4].

Difficulties in recognizing the existence of microcalcification can be caused by many things, but one of them caused by the process of digitization [2]. This digitization process may cause degradation such as noisy and blurry. Image enhancement technique believed to produce image with better quality [5].

Therefore, this study aims to perform image enhancement in mammography image in recognizing microcalcification. Transformation to negative image and histogram equalization in this study used to process the original mammography image. At initial steps original mammographic image load to the application, then secondly process the image use as input to negative image techniques this technique suited when the dark

region dominant in the image[6], final step histogram equalization used to redistributed the pixel value to get optimal value [2].

II. LITERATURE REVIEW

This section described recent studies and basic image enhancement theory as follows:

A. Recent studies

Microcalcification has characteristics such as normal tissue, to distinguish it required segmentation techniques [3], segmentation process is a technique that aims to distinguish observation areas visually. However, the visual quality of the image is influenced by the density of the observation object[2]. Many techniques have been proposed in recognizing microcalcification [1], [4]. Lots of method used, among them equalization histogram can be used to sharpness improvement [7], then negative image fits when the dark region become the dominant feature [6].

B. Digital image

Digital image can be defined as a two dimensional function $f(x, y)$, where x and y are spatial coordinates and the amplitude of f in any coordinate pair (x, y) is called the gray level of the image at that point [6]. Digitized image as shown in Fig 1.

$$f(x, y) = \begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0, N-1) \\ f(1,0) & f(1,1) & \dots & f(1, N-1) \\ \vdots & \vdots & \ddots & \vdots \\ f(M-1,0) & f(M-1,1) & \dots & f(M-1, N-1) \end{bmatrix}$$

Fig 1 digitized image

C. Negative image

The transformation of the original image into a negative image is required with conditions if the dark areas become the dominant [6]. Transformation to negative image:

$$Gray_{New} = 255 - Gray_{Old} \quad (1)$$

This operation produced negative image [2]. $Gray_{New}$ obtain by subtracting $Gray_{Old}$ with value 255.

D. Histogram equalization

This technique will redistributed pixel value to obtain optimal result [8]

$$w = \frac{C_w Th}{n_x n_y} \quad (2)$$

Where:

w = histogram equalization
 c_w = histogram cummulative
 t_h = threshold (default: 256)
 $n_x - n_y$ = image dimension

III. RESEARCH METODOLOGY

This section described steps involved in this studies as follow:

A. proposed method

This part described sistematically approach as shown in Fig 2.

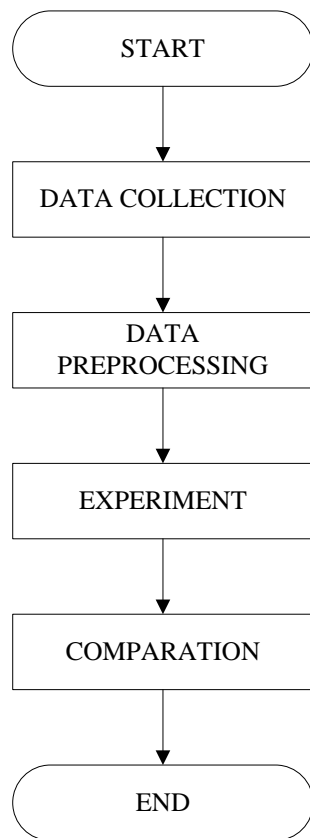


Fig 2 proposed method

- Data collection
Mammographic image obtain from Mammographic Image Analysis Society (MIAS) from <http://peipa.essex.ac.uk/info/mias.html>. image group by normal and positive with cancer. Data format in .PGM (portable gray map). This studies required 8 image and group by normal and cancer positive.
- Data preprocessing

Prepared the working directory for data saving, convert to jpg to make the image dimension same.

- Experiment
.pgm format convert to jpg and transform to negative image and run histogram equalization.
- Comparation
The final result of proposed method compared with the original image that already convert to .jpg format.

IV. RESULT AND DISCUSSION

This section describe the result of the studies of image enhancement on mammographic image as follows:

A. image group

Image used in this studies obtain from Mammographic Image Analysis Society (MIAS) from <http://peipa.essex.ac.uk/info/mias.html> and group as shown Tab 1.

TABLE I. MAMMOGRAPHIC IMAGE

| CLASS | ABNORMALITY | CHAR | SAMPLE |
|-----------|-----------------------------------|---------------------|--------|
| NORM | | FATTY (F) | MDB006 |
| | | FATTY-GLANDULAR (G) | MDB007 |
| | | DENSE-GLANDULAR (D) | MDB003 |
| MALIGNANT | MICROCALCIFICATION | FATTY (F) | MDB231 |
| | | FATTY-GLANDULAR (G) | MDB209 |
| | | DENSE-GLANDULAR (D) | MDB239 |
| | WELL-DEFINED CIRCUMSCRIBED MASSES | FATTY (F) | MDB028 |
| | | FATTY-GLANDULAR (G) | MDB270 |

This table described mammographic image sample group by normal mammographic breast image and cancer breast image. Image enhancement applied to this eight sample image with fatty, fatty-glandular and dense-glandular.

B. image processing

Image format used in this studies is .pgm (portable gray map) with image dimension 1024 x 1024. When those image load in Octave the dimension change to 1200 x 898. There for next step is convert the .pgm image format to .jog image format. Format image transformation taken for MSE (Means Square Error) and PSNR (Peak Signal Noise Ratio) calculation. Different dimension makes the MSE and PSNR calculation failed.

C. Experiment

Image with .jpg format load to the application and used as input for negative process then process to histogram equalization. When two process done. The result compared with the image with .jpg format. The image enhancement process as show in Fig 3. Image processed to image negative develop by Integrated Development Environment GNU Octave, instruction to transformed the .jpg formatted to image negative shown as Fig 4 and instruction to run histogram equalization shown as Fig 5

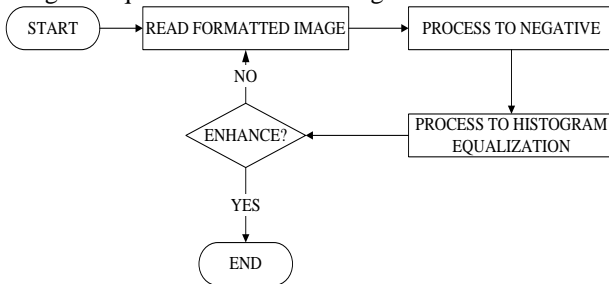


Fig 3 experiment

Fig 3 described the whole process in this studies. At initial steps, mammographic image with .pgm convert to .jpg format. This process need to be done since if the dimension of the image different the MSE and PSNR can't calculated.

```

1 function Negative (hObject, eventdata,
NegCitra)
2     [fname, fpath] = uigetfile();
3     i = imread(fullfile(fpath,
fname));
4     axes(NegCitra);
5     negatif = 255 - i;
6     imshow(negatif, []);
7 end
  
```

Fig 4 image negative instruction

Definition per line as follows:

- Line 1 and 7 : user-defined function
- Line 2 : dialog box *function*
- Line 3 : read the data from line 2.
- Line 4 : axes to display the image
- Line 5 : image negative function
- Line 6 : function to display the image

After the process of image negative done, continued to process this histogram equalization.

```

1 function Histeq (hObject, eventdata,
Histeq)
2     [fname, fpath] = uigetfile();
3     i = imread(fullfile(fpath, fname));
4     negatif=255-i-i;
5     j = histeq(negatif, 256);
6     axes(Histeq);
7     imshow(j, []);
8 end
  
```

Fig 5 histeq function

Definition per line as follows:

- Line 1 and 8 : user-defined function
- Line 2 : dialog box
- Line 3 : read data
- Line 4 : image negative function
- Line 5 : histeq threshold 256
- Line 6 : axes
- Line 7 : display the image

D. MSE and PSNR

The function of MSE (Means Square Error) and PSNR (Peak Signal Noise Ratio) is a common parameter used as an indicator in comparing the similarity of the two images (initial image and processing image). The use of both functions in this study basically aims as a measuring tool and / or to validate the level of similarity. The benefits of using these two functions as an alternative when encountering difficulties to finding experts in the field of image processing and cancer experts. Code to find the MSE and PSNR as shown in Fig 6:

```

1 img=imread();
2 img_result=imread();
3 [row, col, ~]=size(img);
4 mse = sum(sum((img-
img_result).^2))/(row*col);
5 psnr = 10*log10(256*256/mse);
6 disp(mse);
7 disp(psnr);
  
```

Fig 6 MSE and PSNR

Definition per line as follows:

- Line 1 : read the image.
- Line 2 : read the result image
- Line 3 : array variable
- Line 4 : MSE.
- Line 5 : PSNR.
- Line 6 : display MSE.
- Line 7 : display PSNR.

MSE and PSNR show in Fig 7.

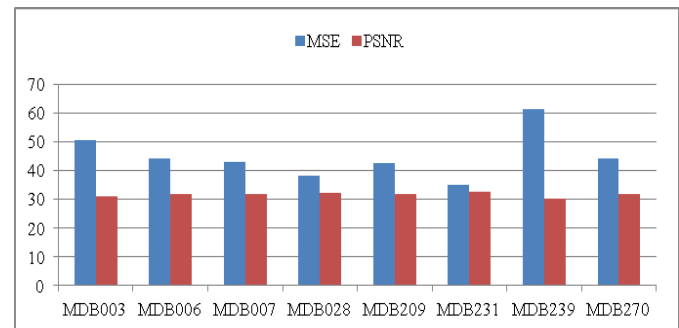


Fig 7 MSE and PSNR

Detail of MSE and PSNR value show in Tab. II. This table defined that only some of mammographic have better visualisation. however the overall process of image enhancement can applied to mammographic image.

TABLE II. MSE and PSNR

| IMAGE | MSE | PSNR |
|--------|-------|--------|
| MDB003 | 50.35 | 31.145 |
| MDB006 | 43.99 | 31.731 |
| MDB007 | 42.9 | 31.84 |
| MDB028 | 37.99 | 32.368 |
| MDB209 | 42.6 | 31.871 |
| MDB231 | 35 | 32.725 |
| MDB239 | 61.44 | 30.281 |
| MDB270 | 44.13 | 31.718 |

E. Comparison

In this section will show the results of the use of image processing using image improvement techniques with the use of negative image function and histogram equalization. Both images are compared to be able to determine the image quality improvement. Improved imagery does not all have good quality images, but there are some image quality improvements. Histogram equalization threshold use 256 as default values. This quality improvement is used to facilitate the process of observation by health personnel. Result of the image enhancement process using negative image and histogram equalization show in Table III.

V. CONCLUSION

Image enhancement process on mammography image can be done, however there are only some image that have improved quality, this affected by threshold usage, which have

important role to get better visualization on mammographic image.

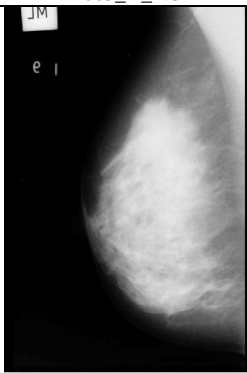
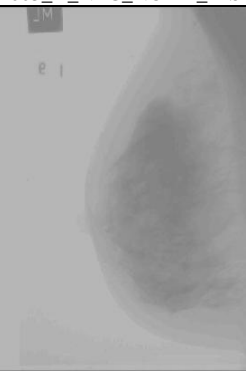

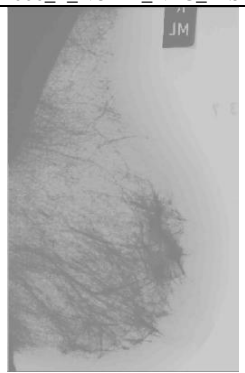
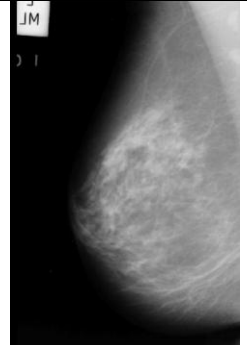
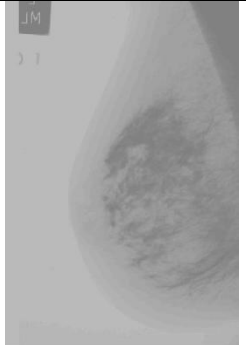
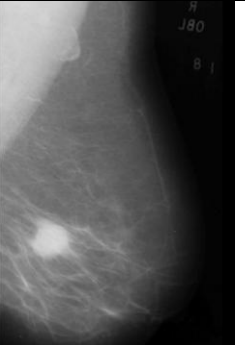
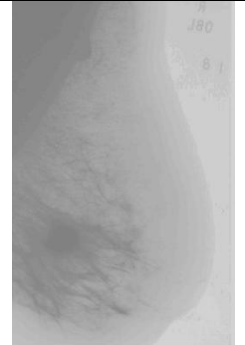
ACKNOWLEDGEMENT

THE RESEARCHER WOULD LIKE TO THANK THE DIRECTORATE OF RESEARCH AND COMMUNITY SERVICE OF DIRECTORATE GENERAL OF STRENGTHENING RESEARCH AND DEVELOPMENT OF RESEARCH, TECHNOLOGY AND HIGHER EDUCATION MINISTRY ACCORDING TO RESEARCH CONTRACT YEAR 2018.

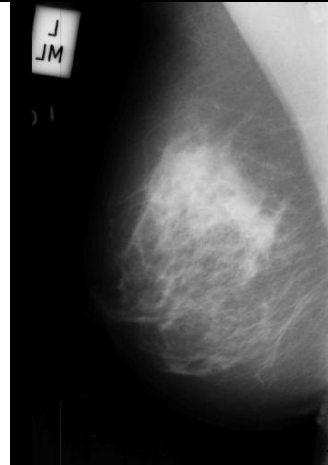
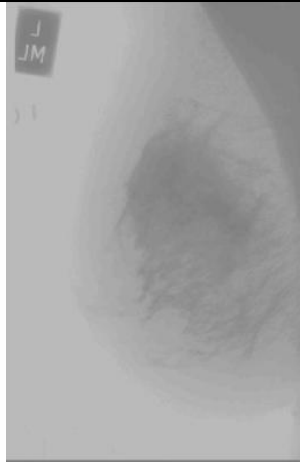
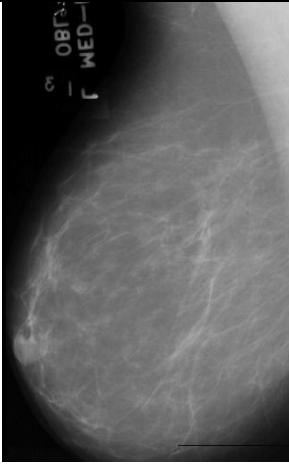
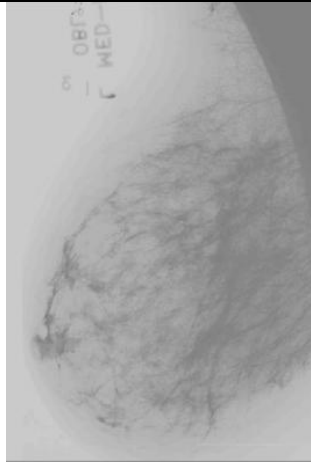
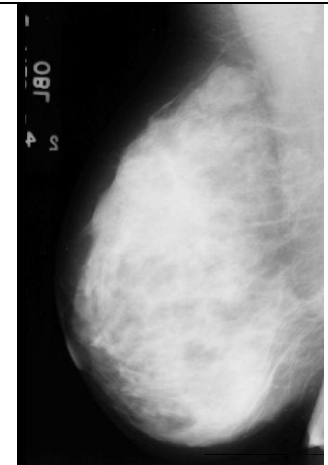
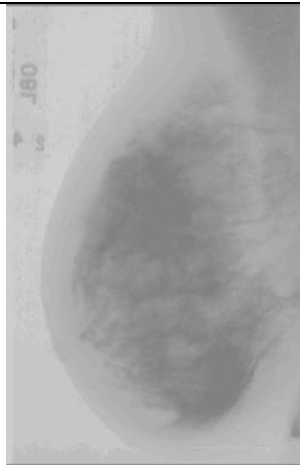
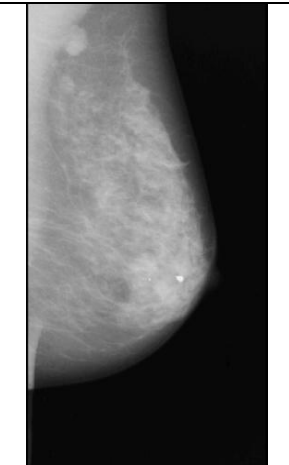
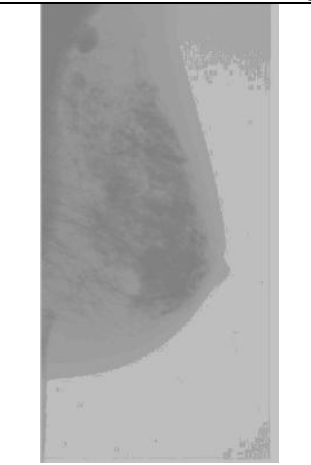
REFERENCES

- [1] N. El Atlas, M. El Aroussi, and M. Wahbi, "Computer-aided breast cancer detection using mammograms: A review," *2014 2nd World Conf. Complex Syst. WCCS 2014*, vol. 6, pp. 626–631, 2014.
- [2] D. Priyawati, "Teknik Pengolahan Citra Digital Berdomain Spasial Untuk Peningkatan Citra Sinar-X," vol. II, pp. 44–50, 2011.
- [3] J. Quintanilla-Dominguez, B. Ojeda-Magañ, M. G. Cortina-Januchs, R. Ruelas, A. Vega-Corona, and D. Andina, "Image segmentation by fuzzy and possibilistic clustering algorithms for the identification of microcalcifications," *Sci. Iran.*, vol. 18, no. 3 D, pp. 580–589, 2011.
- [4] T. Nurhayati and B. Destyningtias, "Identifikasi Kanker Payudara dengan Thermal," *Pros. Semin. Nas. Sains dan Teknol.*, no. 1, pp. 75–79, 2010.
- [5] J. Mohanalin and M. Beena Mol, "A new wavelet algorithm to enhance and detect microcalcifications," *Signal Processing*, vol. 105, pp. 438–448, 2014.
- [6] E. Prasetyo, *Pengolahan Citra Digital dan Aplikasinya menggunakan Matlab*. Penerbit Andi Yogyakarta, 2011.
- [7] C. Y. Wong, G. Jiang, M. A. Rahman, S. Liu, S. C. F. Lin, N. Kwok, H. Shi, Y. H. Yu, and T. Wu, "Histogram equalization and optimal profile compression based approach for colour image enhancement," *J. Vis. Commun. Image Represent.*, vol. 38, pp. 802–813, 2016.
- [8] J. C. Russ, *The Image Processing Handbook*. CRC Press, 2011.

TABLE III. RESULT COMPARATION

| MDB003_D_NORM | MDB003_D_NEG_NORM_HISTEQ | MDB006_F_NORM | MDB006_F_NORM_NEG_HISTEQ |
|---|---|--|---|
|  |  |  |  |
| MDB007_G_NORM | MDB007_G_NORM_NEG_HISTEQ | MDB028_F_CIRC | MDB028_F_CIRC_NEG_NORM_HISTEQ |
|  |  |  |  |

Continued...

| | | | |
|--|--|---|--|
| MDB209_G_CALC | MDB209_G_CALC_NEG_HISTEQ | MDB231_F_CALC | MDB231_F_CALC_NEG_HISTEQ |
|  |  |  |  |
| MDB239_D_CALC | MDB239_D_CALC_NEG_HISTEQ | MDB270_G_CIRC | MDB270_G_CIRC_NEG_HISTEQ |
|  |  |  |  |

Improved K-mean Clustering Algorithm for Prediction Analysis using Classification Technique in Data Mining

Swati Vinodani

Galgotia college of engineering

Swativinodani2015@gmail.com

Aatif jamshed

Assistant professor

Galgotia college of engineering

09.aatif@gmail.com

Dr. Pramod Kumar

Director

Tula's Institute of Technology

drpramoderp@live.com

Abstract- Data mining is utilized to manage huge measure of information which are put in the data ware houses and databases, to discover required information and data. Numerous data mining systems have been proposed, for example, association rules, decision trees, neural systems, clustering, and so on. It has turned into the purpose of consideration from numerous years. A re-known amongst the available data mining strategies is clustering of the dataset. It is the most effective data mining method. It groups the dataset in number of clusters based on certain guidelines that are predefined. It is dependable to discover the connection between the distinctive characteristics of data.

In k-mean clustering algorithm, the function is being selected on the basis of the relevancy of the function for predicting the data and also the Euclidian distance between the centroid of any cluster and the data objects outside the cluster is being computed for the clustering the data points. In this work, author enhanced the Euclidian distance formula to increase the cluster quality.

The problem of accuracy and redundancy of the dissimilar points in the clusters remains in the improved k-means for which new enhanced approach is been proposed which uses the similarity function for checking the similarity level of the point before including it to the cluster.

Keywords: Data Mining, Clustering, Classification, Dataset, k-means, Similarity, Centroid, Data objects, Density.

I. INTRODUCTION

Data mining, the acquiring of required data from huge databases, is an intensive new development in the field with high potential to enable organizations to center around the critical data in their dataset. Different methodologies of data mining tries to extract the pattern on the basis of the future requirement from the dataset and do the job to make the organization to have the active or proactive, learning based kind of work. The different works of the data mining makes the organization to answer the issues related to the dataset which generally are tiring work and also sometime are just impossible kind of work. Some of the concealed cases are being considered so as to scour the database and makes the prescient data available which actually was forbidden by the works done previously [1].

Most organizations officially gather's and refine huge amounts of information. Data mining procedures can work quickly on existing tools and equipment to uplift the advantages of available data resources and can coordinate with new items and frameworks as they are made on-line. At the point when runned on superior customer/server or parallel preparing PCs, data mining devices can manage large databases to convey answers to inquiries, for example, "Which customers are well on the way to react to my next promotion-based mailing, and why?"

Clustering is a data mining procedure that makes noteworthy or supportive group of substance that have comparative element utilizing mechanical strategy. Divergent from order, Clustering procedure likewise characterizes the classes and place data points in them, as in classification objects are considered into predefined classes. For instance, in forecast of coronary illness by utilizing Clustering, get group or express that rundown of patients which have same hazard factor. This influences the split rundown of patients with high blood to sugar and related hazard factor n so on.

The idea of a "cluster" can't be accurately characterized, which is one reason why there are such huge numbers of clustering algorithms. [3] There is a shared factor: a groups of information objects. As, distinctive analysts utilizes diverse cluster models, and for every one of these cluster models again extraordinary techniques can be given. The thought of a cluster, as found by various techniques, shifts fundamentally in its properties.

A. *Clustering Methods*

Clustering methods can be classified into the following categories:

- Partitioning Method
- Hierarchical Method
- Density-based Method
- Grid-Based Method
- Model-Based Method
- Constraint-based Method

B. Data Mining Used in Various Applications [2]

- Business Intelligence
- Sports
- Analyze Students Performance
- Telecommunication Industry
- Retail Industry

C. Issues and Challenges:

- Dissimilarity of the data in the clusters,
- Clustering accuracy,
- To produce fixed and appropriate cluster centers,

II. PROBLEM STATEMENT

In our work we extend the work done by Arpit et al. (2017) [4], “**Improved K-mean Clustering Algorithm for Prediction Analysis using Classification Technique in Data Mining**”, k-means algorithm is being used for database clustering in which the centroid is calculated and then on the basis of the Euclidian distance the data objects from the dataset are then grouped to form of cluster of similar type of clusters. As the data object inclusion in the cluster is being using the Euclidian distance hence in some of the cases the methodology fails to show the accuracy of the clusters in the terms of similarity of the data objects in any clusters and also the major disadvantage of the k-means is the static definition of the threshold for the number of clusters. The major disadvantage of k-means algorithm is that the number of cluster for any dataset are needed to be predefine and just because of which some of the points or objects inside the dataset remains un-clustered.

III. LITERATURE REVIEW

[4] The k-mean clustering techniques is utilized to group the comparative type of information for prediction analysis. In k-mean clustering techniques, probability of the most relevant function is figured and utilizing Euclidian distance equation the objects are grouped. In this work, we will improve the Euclidian distance equation to expand the cluster quality. The upgrade will be based on normalization. In the improvement two new highlights will be included. The primary point is to compute ordinary distance measurements based on normalization. In second point the objects will be grouped based on larger part voting. The proposed strategy will be executed in MATLAB.

[5] K-Means or Hard C-Means clustering is essentially a dividing strategy connected to investigate information and considers perceptions of the information as data point in view of areas and separation between different information points. Partitioning the items into totally unrelated groups (K) is finished by it in such a way, to the point that articles inside each cluster stay as close as possible to each other, and also far from objects in different groups.

Each cluster is represented by its inside point i.e. centroid. The distance utilized as a part of clustering in the vast majority of the circumstances don't really speak to the spatial separations. By and large, the main answer for the issue of finding worldwide least is comprehensive decision of beginning stages. Yet, utilization of a few copies with random beginning stage prompts an answer.

[6] Bezdek presented Fuzzy C-Means clustering strategy in 1981, stretch out from Hard C-Mean clustering technique. FCM is an unsupervised clustering technique that is connected to extensive variety of issues associated with feature investigation, clustering and classifier design. FCM is broadly connected in agrarian engineering, chemistry, geography, image investigation, medical analysis, shape analysis and target acknowledgment [7].

With the improvement of the fuzzy hypothesis, the FCM clustering techniques which is really in relies on Ruspini Fuzzy grouping hypothesis proposed in 1980's. This technique is utilized for analysis of distance between different information objects. The clusters are shaped by the distance between information objects and the cluster center are framed for each cluster.

[8] The DBSCAN techniques was first presented by Ester, et al. [Ester1996], and depends on a density-based idea of clusters. Clusters are recognized by considering the thickness of data points. Areas with a high-density objects shows the presence of clusters though locales with a low density of points demonstrate clusters of noise or clusters of exceptions. This technique is especially suited to manage huge datasets, with noise, and can distinguish cluster with various sizes and shapes.

The key thought of the DBSCAN techniques is that, for each object of a cluster, the area of a given radius needs to contain no less than a base number of data points, that is, the density in the area needs to surpass some predefined limit.

[9] The SNN technique [Ertoz2003], as DBSCAN, is a thickness-based clustering technique. The fundamental distinction between this technique and DBSCAN is that it characterizes the comparability between points by considering at the quantity of closest neighbors that two points share. Utilizing this similarity measure in the SNN technique, the density is characterized as the aggregate of the similarity of the closest neighbors of a point. Points with high density move toward becoming center points, while points with low density speak to noise points. All leftover points that are emphatically comparative a particular center points will speak to another clusters.

[10] DENCLUE (Density based clustering) utilizes two primary ideas i.e. impact and density function. Impact of every data point can be displayed as numerical function. The subsequent function is called Influence Function. Impact work defines the effect of information point inside its neighborhood. Second factor is Density function which is total of impact of all information points. DENCLUE characterizes two kinds of clusters i.e. defined characterized and multi center characterized groups. $y \in F$ is an impact function of the information objects.

Which is characterized as far as an essential impact function F , $F(x) = -F(x, y)$. The density function might be characterized as the whole of the impact elements of all information points. DENCLUE is additionally used to sum

up other clustering techniques like Density based grouping, segment-based clustering, hierarchical clustering. DBSCAN is a case of density based clustering and square wave impact function is utilized.

[11] Fundamentally, DBCLASD is an incremental approach. DBCLASD depends on the assumption that the points inside a cluster are grouped consistently. DBCLASD progressively decides the best possible number and state of clusters for a database without requiring any information parameters [12]. A random point is provided to a cluster which is then prepared incrementally without thinking about the cluster.

In DBCLASD, a cluster may be defined by three properties shown below:

- 1) Expected Distribution condition
- 2) Optimality Condition
- 3) Connectivity Condition

IV. METHODOLOGY

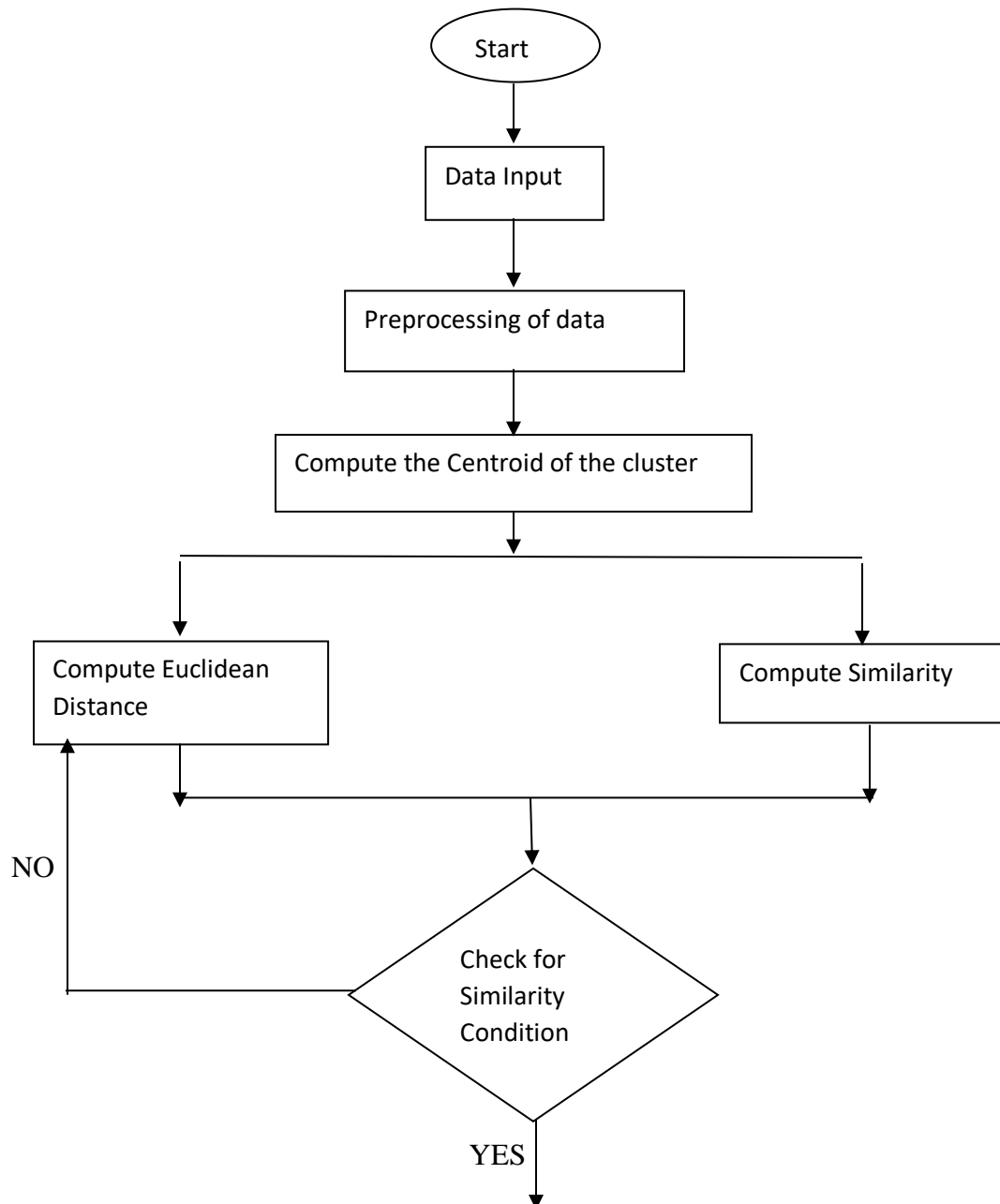
The below section of the paper describes the proposed work “*Enhanced K-means Clustering using Euclidian distance and Similarity Function*”. Some meaningful and required results are being analyzed by the data mining tools available. The facility of the data mining is being used in many applications in the real time for example customer retention, education system, production control, healthcare, market basket analysis, manufacturing engineering, scientific discovery and decision making etc.

In clustering the data objects available in the dataset is being grouped together such like that the data objects with similar properties are together. In the case when the number of clusters in the dataset are less then the finer details of the dataset might be missed for better representation. The clusters are being used for modeling the data. Data modeling puts clustering in a historical perspective rooted in mathematics, statistics, and numerical analysis. As of taking as per the machine learning the data mining is all about the hidden patterns, and also is unsupervised kind of learning and sort of data concept is being shown by the representation using the data mining technique. When the practical consideration of the data mining is being taken into account then there exist an unforgettable role of the data mining tools in many applications like exploration of the scientific information, mining of text and information retrieval, applications based on spatial datasets, analysis of internet, CRM, marketing, medical diagnostics, computational biology, and many others.

In this work we have extended the Arpit Bansal *et al.* (2017), “Improved K-mean Clustering Algorithm for Prediction Analysis using Classification Technique in Data Mining” The k-means methodology is being used for the prediction of the data which are very much similar to each other. A function is being selected on the basis of the relevancy of the function and along with the Euclidian distance is being used for clustering the data points. The enhancement in the k-means methodology is being done on the basis of the normalization of the data. In the enhancement two new features will be added. The first point is to calculate normal distance metrics on the basis of normalization. The selection is being done on the basis of the majority voting which is considered as the second stage in the data mining process.

The problem of accuracy and redundancy of the dissimilar points in the clusters remains in the improved k-means for which new enhanced approach is been proposed which uses the similarity function for checking the similarity level of the point before including it to the cluster.

A. Proposed Architecture



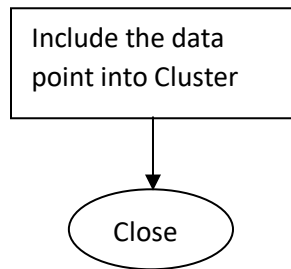


Figure 1. Processing Architecture of the proposed technique.

Step by step execution of the proposed methodology (Enhanced K-means Algorithm):-

Step 1:- Data is being generated from user,

Step 2:- Preprocessing of data is being done like data cleaning for noise removal, data integration and data visualization,

Step 3:- Enhanced k-means is applied over the data:

- Centroids of the clusters is being computed,
- Euclidian distance between centroid and other point is computed using the mathematical formula discussed in step 1 of k-means algorithm,
- The point selected is being checked for similarity before including it to the cluster using the similarity function,

$$SimMsr(C_i) = \sum_{C_{C_i}, C_{p_j} \in C} SimFunc(C_{C_i}, C_{p_j}) \quad (1)$$

Where, C_{C_i} is the centroid of the i_{th} cluster,

C_{p_j} is the cluster point which is to be included,

C is the total clusters.

Step 4:- Check the data point for similarity if the similarity of the nearest point is high then go to step 5 else consider the next closest point for similarity checking.

Step 5:- The point is being included in the clusters if it appropriately defines the cluster similarity.

B. Concept of Similarity

It is natural to ask what kind of standards we should use to determine the closeness, or how to measure the dissimilarity or similarity between a pair of objects, an object and a cluster, or a pair of clusters. This segment of the paper talks about the concept of similarity and also about the hierarchical clustering and the proximity between the clusters is

being estimated. The single cluster is being represented by the prototype so that the cluster can be processed further just similar to any other object present in the dataset.

A data object is described by a set of features, usually represented as a multidimensional vector. The features may be quantitative or qualitative, continuous or binary, nominal or ordinal, which estimates the methodology. A distance or dissimilarity function on a data set is represented to fulfill the following conditions.

Likewise, a similarity function is defined to satisfy the conditions in the following.

- 1- Symmetry:- $S(x_i, x_j) = S(x_j, x_i)$
- 2- Positivity:- $0 \leq S(x_i, x_j) \leq 1$ for all x_i and x_j
- 3- $S(x_i, x_j) = 1$ if $x_i = x_j$ and it is termed as similarity matrix.

For a data set with input N patterns, we can define an $N \times N$ symmetric matrix, called proximity matrix, whose $(i, j)^{th}$ element represents the similarity or dissimilarity measure for the i^{th} and j^{th} patterns ($i, j = 1, \dots, N$).

Typically, the function for distance estimation which is being utilized for estimating the features which are just continuous, and also the similarity estimation is considered as the qualitative variables. The process of method selection is just issue dependent. For binary features, a similarity measure is commonly used (dissimilarity measures can be estimated by $D_{ij} = 1 - S_{ij}$). Suppose we use two binary subscripts to count features in two objects. n_{00} and n_{11} represent the number of simultaneous absence or presence of features in two objects, n_{01} and n_{10} and count the features present only in one object. Then two types of commonly used similarity measures for data objects x_i and x_j are illustrated in the following.

$$S_{i,j} = \frac{n_{00} + n_{11}}{n_{11} + n_{00} + w(n_{10} + n_{01})} \quad (2)$$

Where,

$W=1$, simple matching coefficient

$W=2$, Rogers and Tanimoto measure.

$W=1/2$, Gower and Legendre measure

The above equation is being used to estimate the similarity in any two data objects. The pairs which remains un-matched are provided with some weights and which decides the similarity of the objects or data objects.

$$S_{i,j} = \frac{n_{11}}{n_{11} + w(n_{10} + n_{01})} \quad (3)$$

Where,

W=1, Jaccard coefficient

W=2, Sokal and Sneath measure.

W=1/2, Gower and Legendre measure

These measures focus on the co-occurrence features while ignoring the effect of co-absence. For nominal features that have more than two states, a simple strategy needs to map them into new binary features, while a more effective method utilizes the matching criterion

$$S_{i,j} = \frac{1}{d} \sum S_{i,j} \quad (4)$$

where

$$S_{i,j} = \begin{cases} 0 & \text{if } i \text{ and } j \text{ do not match} \\ 1 & \text{if } i \text{ and } j \text{ match} \end{cases}$$

The proposed methodology starts from the point of getting data from the user as the dataset to be used is dynamic in nature hence data is like changing very frequently. Data pre-processing stands for noise removal and other related steps which can be used so as improve the quality of the data or representation of data. The centroid of the clusters is then computed for each cluster's as the number of clusters are predefined for the particular dataset.

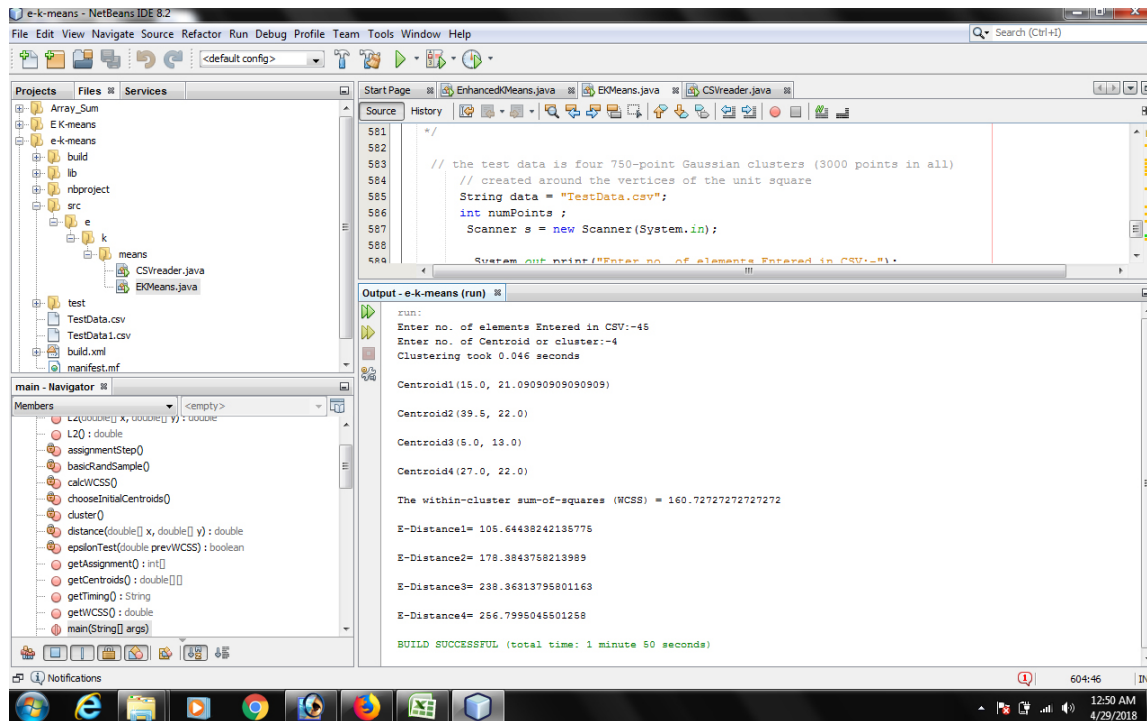
For every centroid the Euclidean distance with the available data point or newly entered data point is being computed and considered in the way of increasing the distance from the centroid. In the very next step the similarity of the data point with the centroid is being evaluated for examining that how much the data point which is to be included look like the centroid.

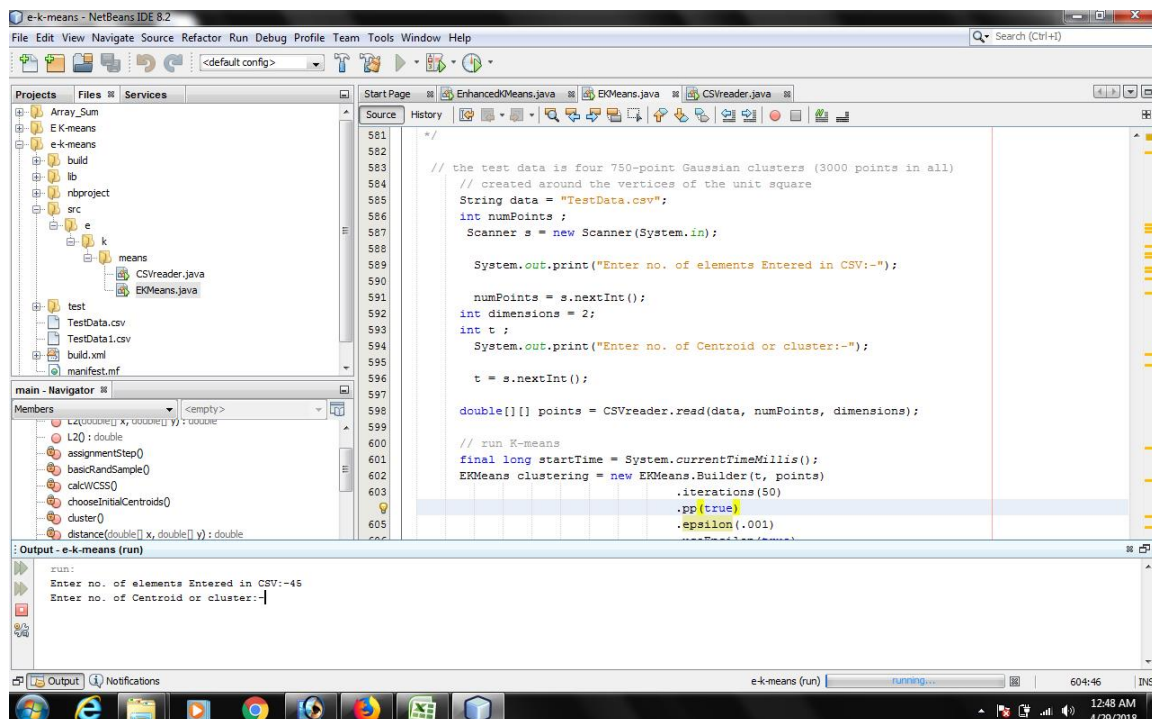
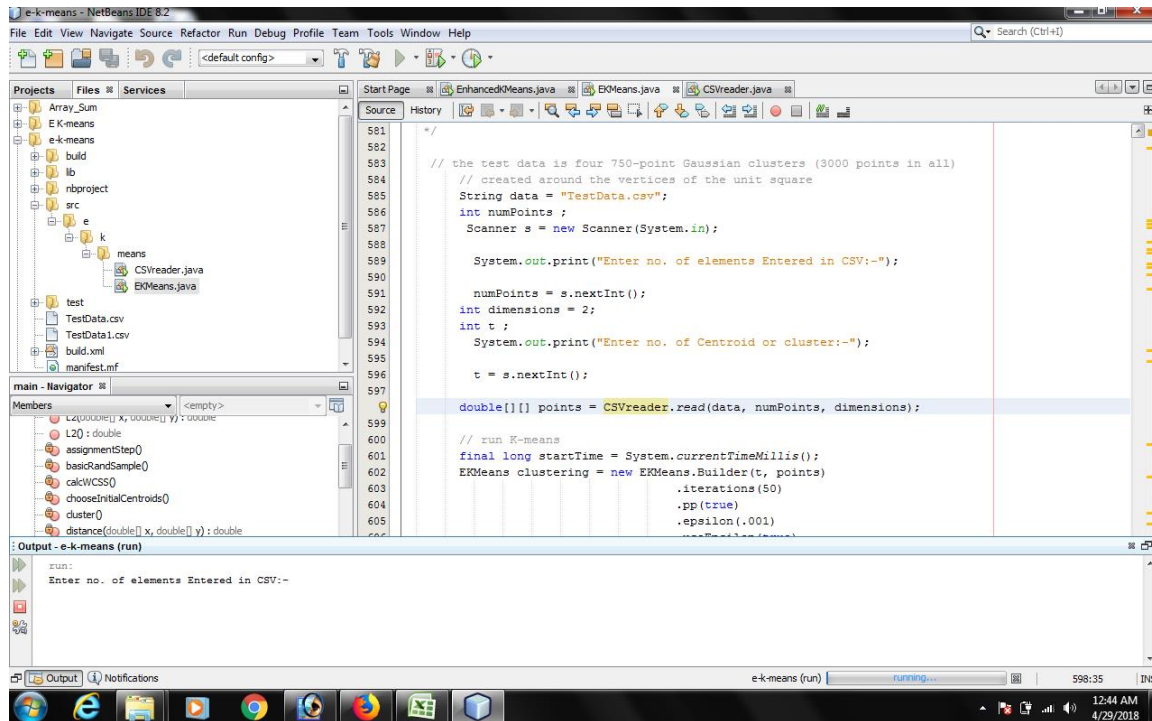
If the similarity of the data point selected on the basis of the shortest distance from the centroid has highest similarity, then it is included in the cluster else the very next data point with respect to the distance is then selected for similarity checking and hence the process keeps on searching till the time the most similar data point is found. As the number of iterations will increase in the overall checking hence will impact the complexity of the overall process but will provide better clusters in which data points included will be of similar properties.

V. ANALYSIS

Both real and the synthetic datasets are being experimented. The synthesized datasets are created with the help of data generator, which follows the basic spirits of well-known IBM synthetic data generator for the generation of clusters. The data size, the number of objects and the average size of the transactions are the important features in the obtained data.

GUI screenshots





The results are represented in the form of clusters of the data points. As the number of objects are increased the proposed methodology outperforms the related techniques for data mining of the large databases in the terms of

the efficiency of the clusters and also it is represented that the proposed work is best suited for mining the real dataset-based applications.

Advantages of proposed methodology

The proposed methodology is advantageous with the other related techniques for mining the large databases in the following aspects:

- The running time is stable for the complete range of number of clusters for the complete synthetic dataset.
- The methodology is more robust as compared to previous others as it have two analysis parameters (Euclidian Distance and Similarity).
- There is a large gap in the execution time when the number of items is increased to an extent; hence out performs other previous techniques for mining the large databases.
- As the execution time in the case of large number of items is efficient hence the proposed methodology is quietly more suitable for the real data mining applications.
- The efficiency of the clusters is higher as the similarity is considered for the inclusion of the data points in any cluster with respect to the evaluation done with the cluster centroid.

VI. CONCLUSION

The paper discuss a novel technique for clustering of the dataset which enhances the k-means clustering which uses similarity of the data objects with each other and also with the centroid of the cluster as well in each iteration. If the similarity of the data objects selected on the basis of the separation from the centroid has highest similarity, then it is included in the cluster else the very next data point with respect to the distance is then selected for similarity checking and hence the process keeps on searching till the time the most similar data point is found. As the number of iterations will increase in the overall checking hence will impact the complexity of the overall process but will provide better clusters in which data points included will be of similar properties.

REFERENCES

1. Berkhin P. (2006), "Survey of clustering data mining techniques in grouping multidimensional data", Springer, pp. 25–71.
2. R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghavan (1998), "Automatic subspace clustering of high dimensional data for data mining applications", in Proc. ACM SIGMOD Int. Conf. Management of Data, pp. 94–105.
3. Abbasi A, Younis M.(2007), "A survey on clustering algorithms for wireless sensor networks", Computer Communications, vol. 30, pp. 14-15.
4. Arpit Bansal et al.(2017), "Improved K-mean Clustering Algorithm for Prediction Analysis using Classification Technique in Data Mining", International Journal of Computer Applications (0975 – 8887) Volume 157 – No 6, pp. 35-40.
5. T. Kanungo and D. M. Mount (2002), "An Efficient K-means Clustering Algorithm: Analysis and Implementation", Pattern Analysis and Machine Intelligence, IEEE Transactions on Pattern Analysis and Machine Intelligence. vol. 24, no. 7.
6. J. C. Bezdek (1981), "Pattern Recognition with Fuzzy Objective Function Algorithms", New York: Plenum Press.
7. Zhang T, Ramakrishna R, Livny M.(1996), "BIRCH: An efficient data clustering method for very large databases", Proceedings of the ACM SIGMOD, International Conference on Management of Data, vol. 25(2), pp. 103–14.

8. Martin Ester, Hans-Peter Kriegel, Jörg Sander and Xiaowei Xu (1996), "A Density- Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise", The Second International Conference on Knowledge Discovery and Data Mining (KDD-96), Portland, Oregon, USA.
9. Levent Ertoz, Michael Steinback, Vipin Kumar (2003), "Finding Clusters of Different Sizes, Shapes, and Density in Noisy, High Dimensional Data", Second SIAM International Conference on Data Mining, San Francisco, CA, USA.
10. Nagpal, Pooja Batra, and Priyanka Ahlawat Mann (2011), "Comparative study of density-based clustering algorithms", International Journal of Computer Applications 27.11, pp. 421-435.
11. XU, X., ESTER, M., KRIEGEL, H.-P., and SANDER, J.1998, "A distribution-based clustering algorithm for mining in large spatial databases", In Proceedings of the 14th ICDE, Orlando, FL, pp. 324-331.
12. Berkhin P. (2006), "Survey of clustering data mining techniques in grouping multidimensional data", Springer, 2006, pp. 25–71.
- 13.** Tian Zhang, Raghu Ramakrishnan, Miron Livny (1996), "BIRCH: An Efficient Data Clustering Method for Very Large Databases", SIGMOD '96 6/96 Montreal, Canada, pp. 103-114.

LECTURERS' ATTITUDES TOWARDS INTEGRATION OF E-LEARNING IN HIGHER EDUCATION. CASE STUDY: UNIVERSITY OF TETOVO

Gazmend Xhaferi
University of Tetovo
Tetovo, Macedonia
gazmend.xhaferi@unite.edu.mk

Rovena Bahiti
University of Tirana
Tirana, Albania
rbahiti@hotmail.com

Abstract— Nowadays E-learning become new way of learning and teaching in higher education. The modern technologies particularly Information and communication technologies, Web 2.0 and the Internet, made higher education no longer limited to the classroom. The purpose of this paper is to investigate lecturers' attitudes toward ICT and integration of E-learning system in higher education. Also the study examine the factors influencing lecturers' attitudes towards ICT and e-learning system. The study was conducted at University of Tetovo, one of the largest public universities of the Republic of Macedonia, where the language of study is the Albanian language. The research developed an extended Technology Acceptance Model (TAM) model for predicting the integration of E-Learning. Statistical analysis was conducted to assess lecturers' attitudes towards integration of e-learning, and to analyses the relationships between their attitudes and their demographic characteristics, perception of usefulness of technology, perception of ease of use of the technology, skills about technology and previous experience and usage the technology that predict the integration of e-learning system. The findings of the study show that there existed positive relationship between these factors and prediction of the integration e-learning. The findings of this study reveal that the lecturers have a positive attitude towards e-learning as well lecturers who are familiar about computer and information and communication technology differ in their attitude towards e-learning when compared to the lecturers who are not familiar with technology. Attitude plays a vital role in using technology as a strong tool for a positive change. Questionnaire was used to collect data from a sample of 49 lecturers from different program studies. Statistical techniques are used for the analyses of data. The findings indicate that lecturers have an important role in prediction of the integration of E-Learning system in University of Tetovo. The reported findings might be of interest to academics, administrators, and decision-makers involved in planning, developing and implementation of e-learning in University of Tetovo and similar universities in developing countries.

KEYWORDS

E-Learning; Technology Acceptance Model (TAM); ICT Perception; ICT competence; ICT Experience; Attitude;

I. INTRODUCTION

Information and communication technology (ICT) has also brought a paradigm shift in the approach towards learning and teaching in educational system. The advancements of Information and multimedia technology, and the use of internet as a new way of teaching, has made a revolutionary changes in the traditional teaching process (Tao et al., 2006). One of the most significant developments in the use of information technology in universities in the last decade has been the integration and use of e-learning systems to support the processes of teaching and learning. E-Learning is a concept derived from the use of information and communication technologies (ICTs) to revise and transform traditional teaching and learning models and practices has evolved in the past decade. Otherwise, e-learning could be seen as a web-based learning tool that utilizes web-based communication, collaboration, knowledge transfer and training to benefit individuals and organizations. It involves the delivery of teaching materials via electronic media, such as Internet, intranets, extranets, satellite broadcasting, audio/video tape, interactive TV, and CD-ROM and it could use Internet technologies to deliver a broad array of solutions that enhance knowledge and performance (Olatubosun, Olusoga & Samuel, 2015).

E-learning has become popular approach in higher education institutions in many countries. Despite the proliferation of e-learning in higher education institutions, the adoption of e-learning still faces a number of obstacles and challenges in some countries. These obstacles and challenges can be summarized as a lack of ICT infrastructure, leadership, training of instructors and learners, as well as e-learning strategy (Khashkhush, 2011).

The idea of integration and adopting e-learning has become widely accepted across higher education in many developed countries, including the USA, UK, most European countries and Australia (Paredes & Correa, 2010; Saowapakpongchai, 2010). Developing countries also appear to adopt e-learning in their higher education to improve and

enhance the education experience. However, the adoption of this technology as a tool for teaching and learning in Macedonia higher education is still in its early stage, where most universities in Macedonia are still struggling to incorporate e-learning into teaching and learning process. In addition, the developing countries often lack the ability to implement advanced educational practices on their own (Andersson & Grönlund, 2009).

Lumumba(2007) recommended that for the implementation of e-Learning in educational institutions to be successful, factors determining the readiness to integrate e-learning system need to be established and dealt with adequately before the implementation process commences. However, successful implementation of e-learning in education relies much on lecturers' attitudes towards it (Avidov-Ungar & Eshet-Alkay 2011; Salmon 2011; Teo 2011; Teo & Ursavas 2012). Among other factors, teacher related variables are the most powerful predictors of technology integration (Becker, 2000).

According to Schiler (2003), personal characteristics of academics staff such as educational level, age, gender, educational experience, experience with the computer for educational purpose and attitude towards ICT can influence the adoption of a technology. Also, several reviews of international literature found that ICT attitudes are influenced by training, knowledge, computer anxiety, computer experience, perceptions of ease of use and usefulness (Buabeng-Andoh, 2012; Fu, 2013; Sab-zian & Gilakjani, 2013). To successfully initiate and implement educational technology in teaching and learning process depends strongly on the lecturers' support and attitudes.

The issue of ICT use in higher education and the factors influencing the integration of E-Learning by lecturers in University of Tetovo have not been extensively investigated before. This study investigate factors that influence lecturers' attitudes towards using ICT in learning and teaching process and the effect of this attitudes on the integration of e-learning at the University of Tetovo. University of Tetovo is one of public university in Macedonia where the teaching and learning language is albanian language. The study uses the Technology Acceptance Model (TAM), which postulates that the subjective norms and perceptions of individuals influence attitudes towards a technology, with attitude as the best predictor of the intention to adopt a technology (Shin & Kim, 2008). This study proposes the use of an extended TAM model, which determines perception about technology, experience and usage of technology and technology skills are primary factors influencing lecturers' attitude and the intention to adopt the technology.

The purpose of the paper is to analyze the factors that affect the integration of e-learning. Concretely will be analyzed among the key factors affecting the attitude about using technology by the academic staff and the integration of e-learning. The model used in this study identifies key variables that can be measured and analyzed to support an empirical assessment of the effect of the variables on the intention to adopt E-Learning system in higher education.

II. THEORETICAL BACKGROUND

In this section have been done literature review about the factors affecting the integration of e-learning by the academic staff. It is also important to identify a theoretical approach to the relationship between academic staffs' attitudes about using ICT and the integration of e-Learning in higher education as a contemporary method in the teaching process. Hall and Khan (2003) describe technology adoption as a consistent process that enables hesitant users to successfully adopt and use technology for a particular purpose. They pointed out that technology adoption occurs when users engage in a series of decisions that are outcomes of comparison of the advantages and disadvantages associated with the use of particular technologies.

Several factors influencing the integration of e-learning into learning and teaching process in higher education have been identified by researchers. According to Schiler (2003), personal characteristics of academic staff such as educational level, age, gender, educational experience, experience with the computer for educational purpose and attitude towards computers can influence the adoption of a technology. Therefore, an understanding of personal characteristics that influence lecturers' adoption and integration of ICT into teaching is relevant. Alazam et al. (2013) identified a close relationship between having technology usage skills and level of technology integration in classroom. Similarly, other studies pointed out that better technology integration into the classroom is dependent on users' level of knowledge and technological skills (Buntat, 2010; Saud et al., 2010). In some studies, the lack of computers and access to them, lagging ICT infrastructural development, cost of training materials, and poor ICT competency skills are identified as significant barriers to technology adoption (Bonsu et al., 2013).

ICT experience has been indicated by many studies to have a significant effect on the behavioral intention of using e-learning system (Park, 2009). Many research studies identified correlations between positive computer experience and positive attitudes, competence and comfort with computers (Papaioannou & Charalambous, 2011; Paris, 2004) and an inverse relationship between computer experience and computer anxiety (Olatoye, 2009). Hence, it is vital to measure lecturers' perceptions about how their computer experience can assist them in accepting and using e-learning.

To successfully integrate and implement educational technology in school's program depends strongly on the lecturers' support and attitudes. Therefore, The attitudes of lecturers towards technology greatly influence their adoption and integration of technology into their teaching and learning process. It is believed that if lecturers perceived technology programs as neither fulfilling their needs nor their students' needs, it is likely that they will not integrate the technology into their teaching and learning. Among the factors that influence successful integration of ICT and e-learning into teaching are lecturers' attitudes and beliefs towards technology (Hew and Brush, 2007; Keengwe and Onchwari, 2008). If lecturers' attitudes are positive toward the use of

educational technology then they can easily provide useful insight about the adoption and integration of ICT into teaching and learning processes. Research has shown that lecturers' attitudes towards technology influence their acceptance of the usefulness of technology and its integration into teaching, Huang & Liaw (2005).

There are various theories of technology acceptance used to appreciate the perceptions of lecturers. One of such models is the technology acceptance model (TAM) developed by Davis (1989 cited in Saade, Nebbe and Tan, 2007; Pituch and Lee, 2006). TAM has been applied to explain or predict individual behaviors across a broad range of end-user computing technology, and user groups (Davis, et al., 1989). The TAM is based on principles derived from psychology, which attempts to understand and measure the "behaviour relevant components of attitudes" and makes possible the understanding of how external stimuli can influence the beliefs, attitudes and behaviour of the individual towards such a thing as technology (Davis, 1993, p. 476).

In this paper, the Technology Acceptance Model (TAM) is used to explain the readiness of lecturers towards using ICT and accept integration of e-learning. Although TAM's ultimate goal is actual usage, it could also be used to explain why individuals may accept or not accept a particular technology such as e-learning (Jung et al., 2008).

Therefore, the model can be effectively applied to analyze specific variables that affect the individual's decision to accept the use of technology. Issues such as the perceived usefulness of technology and ease of use of technology are considered as essential elements for understanding the acceptance of technology. These two main factors facilitate the decision of the individual or group and explain how the integration of technology will take place. Schneberger, Amoroso and Durfee (2008) noted that this model provides a method for understanding the process by which technology is used by the individual. By examining specific factors related to the perceived usefulness and perceived ease of use, this model provides important insights regarding the development of attitudes and behaviours towards technology. Perceived usefulness, in this case, is defined as "the extent to which a person believes that using a technology will enhance her/his productivity". Perceived ease of use is "the extent to which a person believes that using a technology will be free of effort" (Schneberger, Amoroso & Durfee, 2008, p. 76).

The construct attitudes postulates that to the extent that lecturers members perceive the technology is easy to use and helpful, they will have positive attitudes about the technology. Therefore, it is important to consider how this model can be used for understanding both ICT and E-Learning adoption in higher education.

The role of lecturers, their attitudes towards use of ICT in learning and teaching process and factors influence these attitudes and how those attitudes can ultimately contribute to the proliferation of E-Learning must be further examined. The relationship between these factors and lecturers' attitudes towards ICT influence each other in the development of technology and E-Learning acceptance to be

integrate in teaching and learning process must also be considered. These relationships should be determined through an investigation of the lecturers characteristics and the specific external variables that influence their attitudes towards ICT including perception about usefulness of technology and perception of ease of use of technology (**Perception about ICT**), technology usage and experience (**ICT Experience**), and level of knowledge and skills in using the technology (**ICT Competence**). The TAM is expected to be relevant in understanding the lecturers' attitude towards use ICT and integration of e-learning in University of Tetovo.

III. RESEARCH MODEL AND HYPOTHESIS

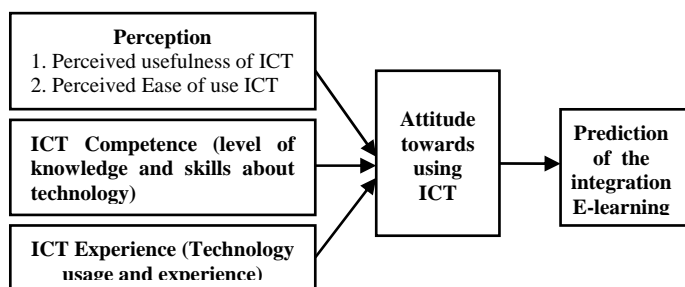
The objective of the research is to investigate the factors that influence the lecturers' attitudes towards using ICT and integration of e-learning system at the University of Tetovo.

The principle of TAM is that people behavioral intention to accept and actually use a certain technology is determined by two constructs namely; perceived usefulness and perceived ease of use. User's attitude and belief as proposed by TAM is perceived to be an important factor which influences the use of new technology. People who have positive attitudes toward information technology will have higher acceptance of the use of the technology in question, compared to people who have negative attitudes toward that technology. Many empirical research (e.g. Davis et al. 1989; Agarwal & Karahanna & Straub, 1999; Venkatesh et al. 2003, 2007) have been carried out and they have shown a support for the favor of TAM.

A. Research model

The research model is presented in Figure 1. However, for the purpose of model development of this research, the TAM model will be expanded including these external variables: perception towards usefulness of technology, perception towards ease of use of technology, technology usage and experience, and level of knowledge and skills about technology, which all have proven to be important factors that influence lecturers behavioral intentions toward adopting a new system.

FIGURE 1. RESEARCH MODEL



It is important to understand how external factors might influence lecturers' attitudes towards ICT, and how those attitudes can ultimately contribute to the prediction of integration of E-Learning by lecturers.

In this research, Attitudes towards using ICT (technology) is considered a dependent variable, which depend from these factors (independent variables): perception towards usefulness of technology, perception towards ease of use of technology, technology usage and experience, and level of knowledge and skills about technology. In order to predict, determine and explain lecturers e-learning integration acceptance, there is a need to understand and assess the extent to which each of the external variables influence the e-learning integration in teaching and learning process. According to Venkatesh (2000), Perceived usefulness is the extent to which a person believes that using a technology will enhance her/his productivity (Venkatesh, 2000) since all else being equal, the less effortful a system is to use, the more using it can increase job performance (Venkatesh and Davis, 2000). Perceived ease of use is the extent to which a person believes that using a technology will be free of effort (Venkatesh, 2000). Teo (2011) and Seif et al. (2012) found direct impact between perceived usefulness and Attitude Towards use in the context acceptance e-learning and factors that affect lecturers and students to use technology. According to Bordbar (2010), lecturers' computer knowledge and skills is a major predictor of integrating ICT in teaching. Evidence suggests that majority of lecturers who reported negative or neutral attitude towards the integration of ICT into teaching and learning processes lacked knowledge and skills that would allow them to make "informed decision" (Bordbar, 2010). Previous usage of technology and experience can be seen as a result of an individual's interaction with the change proponents, or the innovation (or similar innovations) being introduced (Lippert & Davis, 2006). Likewise, in the adoption of innovations, prior experience has been found to have a relative significant influence on the determinants of adoption of new innovations (Taylor & Todd, 1995; Fishbein & Ajzen, 1975). Similarly, a positive or negative experience of the use of technology in education will have an influence on the prediction about integration of E-Learning.

To successfully initiate and implement educational technology in educational program depends strongly on the lecturers' support and attitudes. Among the factors that influence successful integration of ICT and E-learning into teaching and learning process are lecturers' attitudes and beliefs towards technology (Hew and Brush, 2007; Keengwe and Onchwari, 2008). If lecturers' attitudes are positive toward the use of educational technology then they can easily provide useful insight about the adoption and integration of E-learning into teaching and learning processes.

TAM is helpful for both prediction and explanation in the sense that through user's internal beliefs and different significant variables, the researcher can identify reasons that lead to adoption or rejection of e-learning and find appropriate corrective measures or explanations for that decision (Davis et al., 2003; Turner, Kitchenham, Brereton, Charters, & Budgen,

2010). The TAM is easy to extend and validate whilst results from applying the extended TAM are often accepted as being accurate predictors of adoption as well as usage (Davis 1989; Legris, Ingham, & Colletette, 2003).

The extended TAM model in this paper will be used to assess the extent of these factors on lecturers' attitudes towards using ICT and predict integration of e-learning system in teaching and learning process in university of Tetovo. Findings like these confirm that the relationship between lecturers' perception and attitude towards using ICT could manifest an indirect but significant influence on the integration or adoption of E-Learning in Macedonia, especially at University of Tetovo.

B. Research hypothesis

The research aims to test the expanded TAM model and to find the extent to which the role of the three factors, namely; perception about ICT, ICT competence (level of knowledge and skills of ICT) and ICT experience (technology usage and experience) play in the adoption of E-Learning.

The study is divided into two phases. The first stage involved taking the variable Attitude as the dependent variable and all other variables as the independent variables. Taking Perception about usefulness of technology and perception of ease of use of technology, technology usage and experience, and level of knowledge and skills in using the technology (ICT) separately as independent variables.

H1. Lecturers Perceptions about ICT have a positive effect upon Attitude towards using ICT and e-learning.

H2. Lecturers ICT Competence have a positive effect upon Attitude towards using ICT and e-learning.

H3. Lecturers ICT experience has a positive effect upon Attitude towards using ICT.

The second stage involved taking the variable E-learning Prediction (shortened for Prediction of Adoption of E-learning) as the dependent variable and all other variables as the independent variables. In this case taking all four factors Perception, Competence, Experience and Attitude.

H4. The perception of lecturers on the technology has a positive impact upon the Prediction of the integration e-learning.

H5. Lecturers ICT Competence has a positive impact upon the Prediction of the integration e-learning.

H6. Lecturers ICT experience has a positive impact upon the Prediction of the integration e-learning.

H7. Lecturers Attitudes towards ICT has a positive impact upon the Prediction of the integration e-learning.

IV. RESEARCH METHODOLOGY

A. Method and procedures

Research reported in this article utilized a survey design and analysed using Statistical Package for the Social Sciences (SPSS). The survey instrument consisted of: demographic characteristics of lecturers, perceived ease of use of technology, usefulness of technology in teaching,

technology/applications used, knowledge and skills about technology, attitudes towards technology (participants attitude towards technology use for teaching and learning ICT) and prediction about integration of e-learning in teaching and learning process. The first section gathered information about the demographic characteristics of students. The rest of the sections had questions on access to computers (four items), prior experience (four items), perceived ease of use (five items), perceive usefulness (five items), attitude towards e-learning (four items) and behavioural intention to use e-learning (four items) (Davis, 1989; Ong, Lai, and Wang, 2004).

The population of this study comprised lecturers at University of Tetovo. The University of Tetovo is one of four public universities in Macedonia and was established on 4 June 1994 as the first Albanian language higher education institution in Macedonia, though not recognized as a state university by the national government until January, 2004. For this research was selected University of Tetovo, since it is the only public University in Republic of Macedonia where language of the study is albanian language and, as well as due to the size and diversity. University of Tetova is the second largest university in Macedonia in terms of students and staff members.

A total of 65 questionnaires were distributed to lecturers, comprising the following faculty and departments (programmes), such as Faculty of Natural Science and Mathematics of study programs: biology, physics, chemistry, mathematics, and informatics, and Faculty of Economics of study programs: Economy and Business, Accounting and Finance and Management and Marketing. Out of the 65 techers who received the questionnaire, 49 completed and returned the questionnaire given an seventy percent (75%) rate of return. The actual sample size used in the study was sufficient to meet this target sample size.

The questionnaire was designed using a 5-point Likert scale. Lecturers were asked to indicate their agreement or disagreement with several statements using a 5-point Likert-type scale ranging from strongly agree to strongly disagree.

V. DATA ANALYSIS AND RESULTS

Data was analysed using Statistical Package for Social Science (SPSS) software. The Study collected data from 65 lecturers. Descriptive statistics such as median, frequency, and percentage are used for analysis. Furthermore, factor analysis was also performed to identify key factors that are likely to influence integration. Analysis of data collected about lecturers is given in Table 1. The desired sample size was 65 but the actual number of lecturers who took part in the study was 49, yielding a high response rate of 75 %. Most of the lecturers identified as male accounted a 30 (61.2 %), (%) whereas 19 (38.8%) were females. Lecturers age range was normally distributed with four category, first category is between age range of 23-30 years, (11, 22.45%); second category within age range of 31-40, (12, 24.5%); third

category between 41-50, (15, 30.6%); and finally those over 50 years (11, 22.45 %). Data on teaching experience revealed that there were 11 (22.5%) lecturers with teaching experience less than 6 years and 13 (26.5%) lecturers with teaching experience from 6 to 10 years. Furthermore, 10 (20.4%) lecturers had 11 to 15 years of teaching experience and the last category had 15 (30.6%) lecturers with more than 15 years of teaching experience. In terms of study program, results demonstrates that 12 (24.5%) lecturers belongs to the Informatics study program, 9 (18.4%) Mathematics, 5(10.2%) Physics, 5(10.2) Chemistry, 4 (8.2%) Biology, 8 (16.3) Marketing and Management and 6 (12.2%) Economy and Business.

When asked about the use of e-learning as a tool for teaching and learning, lecturers who have never used e-learning exceeded those who have used, with percentages of 63, compared to 37.

TABLE1. LECTURERS' DEMOGRAPHIC CHARACTERISTICS

| Charachteristics | Category | Frequency & Percentage in the Study | |
|--|--------------------------|-------------------------------------|-------|
| | | N | % |
| Gender | Male | 30 | 61.2 |
| | Female | 19 | 38.8 |
| Study program | Informatics | 12 | 24.5 |
| | Mathematics | 9 | 18.4 |
| | Physics | 5 | 10.2 |
| | Chemistry | 5 | 10.2 |
| | Biology | 4 | 8.2 |
| | Marketing and Management | 8 | 16.3 |
| | Economy and Business | 6 | 12.2 |
| Teaching Experience | 0 - 5 years | 11 | 22.5 |
| | 6 - 10 years | 13 | 26.5 |
| | 11 - 15 years | 10 | 20.4 |
| | Over 15 years | 15 | 30.6 |
| Age | 23 - 30 years | 11 | 22.45 |
| | 31 - 40 years | 12 | 24.50 |
| | 41 - 50 years | 15 | 30.60 |
| | Over 50 years | 11 | 22.45 |
| Use E-learning as Learning Tool | Yes | 18 | 0.37 |
| | No | 31 | 0.63 |
| Total No. of techers | | 49 | 100 |

A. Factor analysis

The data was analysed using the SPSS version 16. The descriptive statistics of the five constructs are shown in Table 2. The standard deviations range from 3.58 and 11.56, and all means above midpoint.

Moreover, Construct validity and reliability have been tested to ensure that the results are reliable and consistent. The reliability analysis measured the internal validity and consistency of items used for each construct. Calculating

Cronbach's alpha coefficient tested the factor reliability. This measures the internal consistency by indicating how a set of items are closely related as a group (Moola and Bisschoff, 2012). Recommended by Dunn-Ranking (2004) that a Cronbach alpha value of 0.7 is acceptable, with a slightly lower value might sometimes be acceptable. Cronbach's alpha values for all factors are above 0.70 (see Table 3) indicating that all measures employed in this study demonstrate a satisfactory internal consistency. Therefore, the survey is considered a reliable measurement instrument.

TABLE 2: DESCRIPTIVE STATISTICS

| Factors | N. of Question | Min | Max | Mean | Std dev. |
|--------------------------|----------------|-------|-------|---------|----------|
| PERCEPTION about ICT | 14 | 17.00 | 70.00 | 54.0000 | 11.56 |
| ICT COMPETENCE | 5 | 9.00 | 25.00 | 18.9388 | 3.58 |
| ICT EXPERIENCE | 7 | 7.00 | 32.00 | 19.5102 | 6.68 |
| ATTITUDE towards ICT | 5 | 9.00 | 25.00 | 18.9184 | 3.79 |
| PREDICTION of E-Learning | 11 | 22.00 | 52.00 | 39.3265 | 5.27 |

TABLE 3: ALPHA COEFFICIENTS FOR CONSTRUCTS WITH MULTIPLE ITEMS

| Construct | Cronbach Alpha | Number of Items |
|--------------------------------------|----------------|-----------------|
| PERCEPTION about ICT | 0.952 | 14 |
| ICT COPETENCE | 0.763 | 5 |
| ICT EXPERIENCE | 0.889 | 7 |
| ATTITUDE towards ICT | 0.906 | 6 |
| PREDICTION of integration E-Learning | 0.724 | 11 |

B. Correlation Analysis

The Pearson correlation coefficients were used to measure the relationships between the variables. Correlation analysis answers the question if there exists association or correlation between the two (or more) variables and to what degree. The correlation coefficients were interpreted by using Davis (1971) descriptors (negligible = 0.00 to 0.09; low =0.10 to 0.29; moderate = 0.30 to 0.49; substantial =0.50 to 0.69; very strong = 0.70 to 1.00). The correlation matrix is presented in Table 4. The coefficient correlation results revealed that higher education lecturers' attitudes towards ICT usage was significantly positive strong correlated with their Perceived usefulness of ICT ($r=.781$, $p=.000$), Perceived ease of use of ICT ($r=.748$, $p=.000$), Perception about ICT ($r=.841$, $p=.000$), ICT Competence ($r=.787$, $p=.000$), ICT Experience($r=.629$, $p=.000$).

TABLE 4. THE CORRELATION MATRIX OF FACTORS

**, Correlation is significant at the 0.01 level (2-tailed).

*, Correlation is significant at the 0.05 level (2-tailed).

| | PERCEPTION about ICT | ICT COMPETENCE | ICT EXPERIENCE | ATTITUDE towards ICT | PREDICTION of integration E-L |
|-------------------------------|----------------------|----------------|----------------|----------------------|-------------------------------|
| PERCEPTION about ICT | | | | | |
| ICT COMPETENCE | .788 ** | | | | |
| ICT EXPERIENCE | .616 ** | .686 ** | | | |
| ATTITUDE towards ICT | .841 ** | .787 ** | .629 ** | | |
| PREDICTION of integration E-L | .327 * | .400 ** | .345 * | .478 ** | |

There was a significantly moderate positive relationship between lecturers' Prediction of integration E-Learning and Perception about ICT ($r=.327$, $p=.022$). The results also show a moderate positive correlation between lecturers' Prediction of integration E-Learning and others others variables, such as ICT Competence ($r=.400$, $p=.004$), ICT Experience($r=.345$, $p=.015$) and Attitude towards using ICT($r=.478$, $p=.001$).

For testing hypothesis is used a linear regression analysis that was undertaken using the dependent variables for integration e-learning using the method enter. Regression analysis answers the question if there is any cause and effect relationship between the dependent variable and two or more independent variables and to what degree and in which direction. The hypotheses are tested by the Statistical Package for Social Sciences (SPSS) software.

From the results in the table 5, the first Hypothesis (H1), Perception about ICT (PER) as an independent variable and Attitudes towards using ICT (ATT) as dependent variable, as can see, PER has significant influence on Attitudes towards using ICT (ATT) ($p=.000$). Hypothesis 1 (H1) proves to be supported and Perception about ICT has influence on the lecturers' Attitudes towards using ICT in teaching process. Hypothesis 2 (H2) was also tested; ICT experience (EXP) was independent and attitude towards use ICT (ATT) was dependent. The results indicate that ICT experience has significant influence on (ATT) ($P=.000$). Therefore, hypothesis 2 (H2) proves to be supported and ICT experience and usage has influence on the lecturers' attitude (ATT). Also, the results of Hypothesis 3 (H3) shows that ICT Competence (CPT) has a significant influence on attitude towards use (ATT) ($p=.000$). Thus, ICT competence significantly influences the attitude of lecturers towards ICT(ATT).

The table summarizes the result of regression used to test hypothesis H4. Perception about ICT(PER) as an independent variable and Prediction of integration E-learning(PEL) as

dependent variable. From the results as seen, Perception about ICT has significant influence in Prediction of integration E-learning ($p=.022$). Therefore, hypothesis 4 (H4) proves to be supported and Perception has influence on the lecturers' Prediction of integration E-learning that is significant at the 0.05 level. Also, the result for testing the hypothesis 5 (H5) shows that ICT experience has significant influence in Prediction of integration E-learning ($p=.015$), that is significant at the 0.05 level. So, hypothesis 5 proves to be supported and ICT Experience has influence on the lecturers' Prediction of integration E-learning.

Regarding Hypothesis 6 (H6), the regression analysis shows that ICT competence significantly influences Prediction of Integration E-learning ($p=0.04$). The results presented for H7 indicate that ICT skills significantly influences Prediction. So, hypothesis 7 (H7) is deemed to be supported. And, finally, result indicate that attitude towards use (ATT) has significant impact on Prediction ($p=0.01$). So, hypothesis 7 proves to be supported and lecturers' Attitudes towards ICT has influence on the Prediction of integration E-learning.

TABLE 5: SUMMARY OF THE HYPOTHESIS TESTING

| Research Hypothesis | Path | Standardized Path Coefficient (Beta) | t-value | Significance | Results Significance (p) | R ² |
|---------------------|---------|--------------------------------------|---------|--------------|--------------------------|----------------|
| H1 | PER→ATT | .841 | 10.649 | .000 | Supported | .707 |
| H2 | EXP→ATT | .629 | 5.546 | .000 | Supported | .396 |
| H3 | CPT→ATT | .787 | 8.743 | .000 | Supported | .619 |
| H4 | PER→PEL | .327 | 2.374 | .022 | Supported | .107 |
| H5 | EXP→PEL | .345 | 2.523 | .015 | Supported | .119 |
| H6 | CPT→PEL | .400 | 2.993 | .004 | Supported | .160 |
| H7 | ATT→PEL | .478 | 3.731 | .001 | Supported | .229 |

In the table 5 are summaries the results obtained from testing the research hypotheses. The results confirmed that there was a statistical correlation between the predicted directions of the research model. Overall, all of seven hypotheses were supported by the collected data.

After the examination of each of the seven hypotheses was made, The Figure 1 shows the results of the analysis for our proposed model, including the Standardized Coefficient (Beta) and Significance (P).

VI. DISCUSSION AND CONCLUSION

The purpose of this study was to analyze the attitudes of lecturers on the value of technologies in the learning process and the factors that influence their decisions to adopt and integrate these technologies into teaching process. The aim of the analysis was to determine the degree to which these three variables influenced Attitude, which was established by

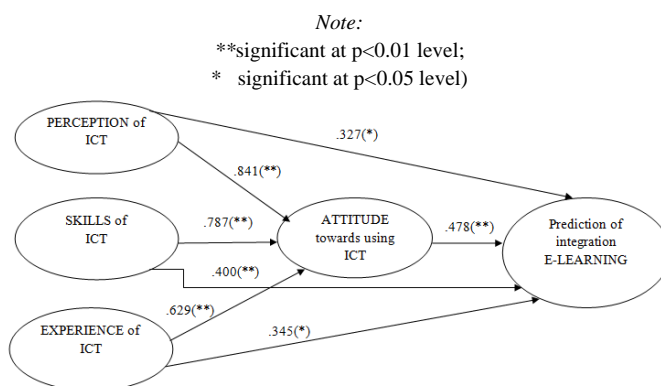
hypotheses h7 as significant for Prediction of integration E-Learning.

From the obtained results we can conclude that all additional factors added to the expanded TAM model have positive lines affecting the goal of integrating an e-learning system. From Figure 2, we note that the factor ICT Competence has a greater impact on lecturers' prediction of e-learning integration, so lecturers who have greater level of skills and knowledge about ICT think positively about integrating e-learning in the teaching process.

Also, findings from the study demonstrate that there was a statistically significant association between variables ICT Experience and prediction of integration e-learning. Thus, it implies that lecturers' experiences in ICT played a significant role in constructing positive attitudes towards integration of e-learning and means that lecturers with higher level of experience and usage of technology in teaching process, the more positive intention towards integration of e-learning in the teaching process. Previous literatures (Albirini, 2006; Pelgrum, 2001) pointed out that lack of computer experience is a main obstacle to teacher's acceptance and adoption of information technologies mainly in developing countries. The results of our research support and extend this finding.

On the other hand, the Perception Factor about ICT that the variable is composed of perceived usefulness of ICT and perceived ease of use of ICT has less impact on the lecturers' prediction about the integration of e-learning in teaching and learning process, who perceive that technology is useful and easy to use in the teaching process. This factor has the least impact compared to other factors analyzed in this study, to lecturers' attitudes about e-learning integration .

FIGURE 2. MODEL OF E-LEARNING TECHNOLOGY INTEGRATION



Finally, the finding of the importance of attitude for the prediction of E-Learning provides further support for the determination that attitude is the key variable in the model as analyzed. The finding of a strong relationship between positive attitude towards ICT and prediction of E-Learning is also similar to the findings of researchers in other nations examining the relationship between attitude and adoption of E-Learning (Inal, Karakus & Cagiltay, 2008). The observations

reported here are consistent with outcomes of similar studies, noting that attitude was a key factor in determining technology adoption (Shiue,2007,. Teo, Lee & Chai, 2008, Teo, 2012)

As mentioned earlier this research identified a number of factors likely to influence the lecturers' attitudes towards integration of e-learning in teaching process. According to Pelgrum (2001), the success of any technology innovation is largely dependent on the skills and knowledge of the educators. Further, Buabeng-Andoh (2012b), mention that success of an educational technology programme in any institution depends on the lecturers' support, and their attitudes as well as the beliefs they hold about the potentials of a particular technology in transforming their teaching practice, and enhancing student learning.

Greater part of lecturers in this study believed and has positive attitude that using ICT in teaching and learning process would significantly contribute to the efficacy and effectiveness of their teaching. Respondents who are already currently employing learning technologies in their teaching said they use them for creativity, to facilitate students' learning, to meet specific

learning objectives, and to perform academic tasks.

In conclusion, findings from this study suggest that lecturers' positive attitude towards ICT and e-learning is essential if SUT higher education institution need to successfully transform its education systems from the current classroom face-to-face methods to e-learning. Lecturers are the key stakeholders of education and their attitudes towards using ICT and also their skills, experience and perception about ICT has a significant impact on prediction of integration of e-learning in learning and teaching process. Identification of attitudes and factors affecting integration of e-learning would provide useful knowledge for education stakeholders and higher institution which can help in planning and increasing effectiveness of the adoption of e-learning in higher education by working out factors, which lead to negative attitudes and strengthening those leading to positive attitudes. The research model and the findings of the study can serve as a model for developing instructional programs to improve ICT skills among lecturers and other stakeholders, which are a prerequisite for influencing attitudes positively towards E-Learning.

REFERENCES

- [1] Tao, Y. H., Yeh, C. R., & Sun, S. I. (2006). Improving training needs assessment processes via the Internet: system design and qualitative study. *Internet Research*, 16(4), 427–49.
- [2] Paredes, Joaquín, & Correa, J-M. (2010). E-learning platforms use and teaching and learning
- [3] models in the European Higher Education Area. Paper presented at the Financial
- [4] Theory and Engineering (ICFTE), 2010 International Conference on.
- [5] Saowapakpongchai, Kultida. (2010). The development of elearning model for higher education
- [6] in Thailand. Paper presented at the Educational and Network Technology (ICENT),
- [7] 2010 International Conference on.
- [8] Andersson, Annika Sophie, & Grönlund, Åke. (2009). A conceptual framework for e-learning in
- [9] developing countries: A critical review of research challenges. *The electronic Journal of information systems in developing Countries*, 38.
- [10] Lumumba, P. (2007). A Survey of Challenges Facing E-learning at Public Secondary Schools: A Case Study of the NEPAD Pilot Project in Schools in Kenya. Unpublished M.ED Project of the Catholic University of East Africa.
- [11] Avidov-Ungar, O., & Eshet-Alkarakay, Y. (2011). Lecturers in a world of change: Lecturers' knowledge and attitudes towards the implementation of innovative technologies in schools. *Interdisciplinary Journal of E-Learning and Learning Objects (IJELLO)*, 7, 291-303.
- [12] Davis, F. D. (1986). *A technology acceptance model for empirically testing new end-user information systems: Theory and results* (Doctoral dissertation, Massachusetts Institute of Technology). Retrieved from <https://dspace.mit.edu/handle/1721.1/15192>
- [13] Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340.
- [14] Salmon, G. (2011). *E-moderating: The key to teaching and learning online* (3rd Ed). London: Routledge. Retrieved from <http://lib.mylibrary.com/Open.aspx?id=336371&src=0>
- [15] Teo, T. (2009). Modelling technology acceptance in education: A study of pre-service lecturers. *Computers and Education*, 52, 302–312.
- [16] Teo, T. (2011). Factors influencing lecturers' intention to use technology: Model development and test. *Computers and Education*, 57(4), 2432-2440.
- [17] Teo, T., Ursavas, O. F., & Bahcekapili, E. (2011). Efficiency of the technology acceptance model to explain pre-service lecturers' intention to use technology: A Turkish study. *Campus-Wide Information Systems*, 28(2), 93-101.
- [18] Teo, T., & Ursavas, O. F. (2012). An assessment of pre-service lecturers' technology acceptance in Turkey: A structural equation modelling approach. *The Asian-Pacific Education Researcher*, 21(1), 191-202.
- [19] Olabode Olatubosun, Fasoranbaku A. Olusoga and Oluwadare A. Samuel (2015). Adoption of eLearning Technology in Nigerian Tertiary Institution of Learning, *British Journal of Applied Science & Technology*,10(2): 1-15, 2015, Article no.BJAST.18434
- [20] Khashkhush, Abdulbasit Salem. (2011). Analysis of factors affecting the implementation of elearning in higher education (HE) in Libya. University of Salford.
- [21] Lumumba, P. (2007). A Survey of Challenges Facing E-learning at Public Secondary Schools: A Case Study of the NEPAD Pilot Project in Schools in Kenya. Unpublished M.ED Project of the Catholic University of East Africa.
- [22] Schiler, J. (2003). Working with ICT: Perceptions of Australian principals, *Journal of Educational Administration*, vol. 41, no. 3, pp. 171-185.
- [23] Buabeng-Andoh C 2012. Factors influencing lecturers' adoption and integration of information and communication technology into teaching: A review of the literature. *International Journal of Education and Development using Information and Communication Technology (IJEDICT)*, 8(1):136-155.
- [24] Fu JS 2013. ICT in education: A critical literature review and its implications. *International Journal of Education and Development using Information and Communication Technology (IJEDICT)*, 9(1):112-125.
- [25] Sabzian F & Gilakjani AP 2013. Lecturers' attitudes about computer technology training, professional development, integration, experience,

- anxiety, and literacy in English language teaching and learning. *International Journal of Applied Science and Technology*, 3(1):67-75. Available at http://www.ijastnet.com/journals/Vol_3_No_1_January_2013/9.pdf. Accessed 25 November 2015.
- [26] Shin, D. H. & Kim, W. Y. (2008). Applying the Technology Acceptance Model and flow theory to cyworld user behavior: implication of the Web2.0 user acceptance. *CyberPsychology & Behavior*, 11(3), pp.378-382.
- [27] Hall, B. H., & Khan, B. (2003). Adoption of new technology (No. w9730). National Bureau of Economic Research. Retrieved on 10 September 2016 from: <http://www.nber.org/papers/w9730>.
- [29] Schiler, J. (2003). Working with ICT: Perceptions of Australian principals, *Journal of Educational Administration*, vol. 41, no. 3, pp. 171-185.
- [30] Alazam, A.-O., Bakar, A., Hamzah, R., & Asmiran, S. (2013). Lecturers' ICT skills and ICT integration in the classroom: the case of vocational and technical lecturers in Malaysia. *Creative Education*, 3(8), 70-76.
- [31] Buntat, Y. (2010). Computer technology application and vocational education: a review of literature and research. *European Journal of Social Sciences*, 14(4), 645-651.
- [32] Bonsu, K., Duodu, A., JA, A., Bonsu, K., Duodu, A., Ansere, J. A., & Djang-Fordjour, K. (2013). The challenges and prospects of ICTs in teaching and learning in sunyani polytechnic, Ghana. *Capa Scientific Journal*, 1(1), 12-17.
- [33] Park, S.Y. (2009). An Analysis of the Technology Acceptance Model in Understanding University Students' Behavioral Intention to Use e-Learning. *Educational Technology, Society*, 12 (3), 150-162.
- [34] Papaioannou, Photos, & Charalambous, Kyriacos. (2011). Principals' attitudes towards ICT and
- [35] their perceptions about the factors that facilitate or inhibit ICT integration in primary schools of Cyprus. *Journal of Information Technology Education: Research*, 10(1), 349-369.
- [36] Paris, Paul G. (2004). E-Learning: A Study on Secondary Students' Attitudes towards Online
- [37] Web Assisted Learning. *International Education Journal*, 5(1), 98-112.
- [38] Olatoye, Rafiu Ademola. (2009). Influence of computer anxiety and knowledge on computer
- [39] utilization of senior secondary school students. *Electronic Journal of Research in Educational Psychology*, 7(3), 1269-1288.
- [40] Hew, K. F., & Brush, T. (2007). Integrating technology into K-12 teaching and learning: current
- [41] knowledge gaps and recommendations for future research. *Educational Technology Research and Development*, vol. 55, pp. 223-253.
- [42] Keengwe, J., & Onchwari, G. (2008). Computer technology integration and student learning:
- [43] Barriers and promise, *Journal of Science Education and Technology*, vol. 17, pp. 560-565.
- [44] Huang, H. M., & Liaw, S. S. (2005). Exploring users' attitudes and intentions toward the Web as a survey tool. *Computers in Human Behavior*, vol. 21, no. 5, pp.729-743.
- [45] Saade, R., Nebede, F & Tan, W (2007). Viability of the "Technology Acceptance Model" in multimedia learning environments: A comparative study. *Interdisciplinary Journal of Knowledge and Learning Objects*, 3, 175-183.
- [46] Pituch, K & Lee, Y-K. (2006). The influence of system characteristics on e-learning use. *Computers and Education*, 47, 222-244.
- [47] Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13, 3, 319-340.
- [48] Davis, F. D., (1993). User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. *International Journal of Man-machine Studies*, 38(3), pp.475-487.
- [49] Jung, M., Loria, K., Mostaghel, R & Saha, P. (2008). E-learning: Investigating University Students' Acceptance of Technology. *European Journal of Open, Distance and E-Learning* [online] http://www.eurodl.org/materials/contrib/2008/Jung_Loria_Mostaghel_Saha.htm
- [50] Schneberger, S. Amoroso, D.L. & Durfee, A. (2008). Factors that influence the performance of computer-based assessments: an extension of the Technology Acceptance Model. *Journal of Computer Information Systems*, 48(2), p.74.
- [51] Karahanna, E., & Straub, D. W. (1999). The psychological origins of perceived usefulness and ease of use. *Information and Management*, 35, 237-250.
- [52] Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27, 425-478.
- [53] Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating perceived behavioral control, computer anxiety and enjoyment into the technology acceptance model. *Information Systems Research*, 11, 342-365.
- [54] Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46, 186-204.
- [55] Teo, Timothy. (2011). Factors influencing lecturers' intention to use technology: Model development and test. *Computers & Education*, 57(4), 2432-2440.
- [56] Seif, Mohammad Hassan, Sarmadi, Mohammad Reza, Ebrahimzadeh, Isa, & Zare, Hossein. (2012). A Model for Predicting Intention to Use E-learning based on Epistemological Beliefs. *Life Science Journal*, 9(2), 926-929.
- [57] Bordbar, F. (2010). English lecturers' attitudes toward computer-assisted language learning. *International Journal of Language Studies*, vol. 4, no. 3, pp. 27-54
- [58] Lippert, S. K., & Davis, M. (2006). A conceptual model integrating trust into planned change activities to enhance technology adoption behavior. *Journal of Information Science*, 32 (5), 434-448.
- [59] Taylor, S., & Todd, P. (1995). Assessing IT Usage: The Role of Prior Experience. *MIS Quarterly*, 19 (4), 561-570.
- [60] Taylor, S., & Todd, P. (1995). Understanding information technology usage: a test of competing models. *Information Systems Research*, 6 (2), 144-176.
- [61] Fishbein, M., & Ajzen, I. (1975). Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research. Reading, MA: Addison-Wesley.
- [62] Keengwe, J., & Onchwari, G. (2008). Computer technology integration and student learning: Barriers and promise, *Journal of Science Education and Technology*, vol. 17, pp. 560-565.
- [63] TURNER, M., KITCHENHAM, B., BRERETON, P., CHARTERS, S. and BUDGEN, D., 2010. Does the technology acceptance model predict actual use? A systematic literature review. *Information and Software Technology*, 52, pp. 463-479.
- [64] VENKATESH, V., MORRIS, M.G., DAVIS, G. B. and DAVIS, F. D., 2003. User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, 27(3), pp. 425-478.
- [65] DAVIS, F. D., 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3), pp. 319-340.
- [66] LEGRIS, P., INGHAM, J. and COLLERETTE, P., 2003. Why do people use information technology? A critical review of the technology acceptance model. *Journal of Information and Management*, 40, pp. 191-204.
- [67] Inal, Y. Karakus, T. & Cagiltay, K., (2008). Turkish high school students. *Turkish Online Journal of Distance Education*, 9(4), p.14.
- [68] Albirini, A. (2006). Lecturers' attitudes toward information and communication technologies: the case of Syrian EFL lecturers. *Computers & Education*, 47 (4), 373-398.
- [69] Pelgrum, W. (2001). Obstacles to the integration of ICT in education: results from a worldwide educational assessment. *Computers & Education*, 37 (2), 163-178.
- [70] Buabeng-Andoh, C. (2012). Factors influencing lecturers' adoption and integration of information and communication technology into teaching: A review of the literature, *International Journal of Education and*

Development using Information and Communication Technology (IJEDICT), Vol. 8, Issue 1, pp. 136-155.

- [71] Teo, T., Lee, B. & Chai, S. (2008). Understanding pre-service lecturers' computer attitudes: applying and extending the Technology Acceptance Model (TAM), *Journal of Computer Assisted Learning*, 24 (2), pp.128-143.
- [72] Teo, T.(2008). Pre-service lecturers' attitudes towards computer use: ASingapore survey. *Australasian Journal of Educational Technology*, vol. 24, no.4, pp. 413-424.
- [73] Teo, T. (2012). Examining the intention to use technology among pre-service lecturers: an integration of the technology acceptance model and theory of planned behavior. *Interactive Learning Environments*, 20(1), 3–18.
- [74] Shiue, Y.-M. (2007). Investigating the sources of lecturers' instructional technology use through the decomposed theory of planned behavior. *Journal of Educational Computing Research*, 36(4), 425–453.

AUTHORS PROFILE

Mr Gazmend Xhaferi is a PhD candidate in the department of Statistical and Applied Informatics at University of Tirana in Tirana, Albania, where he is currently undertaking a research project entitled "Integration and impact of a system for organizing e-learning in the higher education. Case study: University of Tetovo". He is Assistant at Faculty of Natural Sciences and Mathematics at University of Tetovo, in Tetovo, Republic of Macedonia. He has published a series of articles on conferences and journals on the field of e-learning.

Prof. Asoc. Dr. Rovena Bahiti is Associate Professor in the department of Statistical and Applied Informatics at University of Tirana in Tirana, Albania. She has more than 20 papers. She has been involved in numerous research and evaluation projects in the field of education, economy and Financial Accounting. She has been for several years consultative at World Bank and other donor projects in the role of expert in financial management of projects. In December 2013 he was appointed Director of the National Computer Security Agency (ALCIRT), holding it until April 2017

Simulation Based Workflow Scheduling for Scientific Applications

Priya Garg, Dr. Anju Bala

*Computer and Science Engineering Department, Thapar University
Patiala 147004, India*

gpiyag@gmail.com

anjubala@thapar.edu

Abstract—The cloud environment offers an appropriate location for the implementation of huge range of scientific applications. However, in the existing workflows the major dispute is to assign the assets to the tasks in a well-organized way so, that it acquires less finishing time and load on every virtual machines will be impartial. To overcome this problem, GA_ MINMIN has been proposed that combines the features of GA and MINMIN scheduling algorithms. This algorithm is fundamentally a three-layer structure where GA is connected on the main level and hereditary calculation was performed for distributing belonging in an advanced way. At second level, the execution request of the assignments was resolved based on their size. This would be finished with the assistance of MIN-MIN. At third level, all the virtual machines have been running in parallel so that task response time will get decreased with more advanced outcomes. The proposed algorithm has been executed on the simulation environment.

I. INTRODUCTION

Cloud computing is a process that aimed at distributing information technology (IT) amenities in which possessions are regained from the web through web-based gears and tenders, as opposite to a straight linking to a server. The requirements for registering and huge stockpiling assets were quickly developing. In this manner, distributed computing gets the consideration because of the superior figuring administrations and offices that are given to the clients as Software as a Service (SaaS), Infrastructure as a Service (IaaS), and Platform as a Service (PaaS) [1]. Logical researches are generally characterized as workflows, where responsibilities were connected according to their data flow and compute dependencies [2]. Different work processes have diverse structures. As here we were chipping away at two logical work processes such as Montage and CyberShake. The assessment of the execution of work process streamlining procedures in genuine frameworks is unpredictable and tedious.

A. Motivation

Logical workflows were formerly measured for parallel implementation of Cloud computing applications such as Mosaic workflow for lunar physics and CyberShake workflow for tremor risks. These workflows consist of large number of tasks. So, we have proposed GA-Min-Min, to reduce their execution and makespan by scheduling these tasks on virtual machines in an optimized form. Researchers are working on

scheduling of workflows for the reduction of cost, load balancing and minimizing the makespan by implementing various scheduling approaches. GA is an effective approach for assigning tasks to the various resources. As, Hereditary calculations deal with the Chromosome, which is encoded rendering of potential arrangements' parameters, rather the parameters themselves. It seeks parallel from a populace of focuses.

B. Our Contribution

- We have proposed a new hybrid approach that is combination of GA and MINMIN algorithm. With this we are trying to balance the load on different VM's and reducing the response time of each task.
- New feature that is introduced by this algorithm is that we have run all the virtual machines in parallel. This effectively reduced waiting time of each task and because of this more than one task will get the virtual machine at a time.
- Our algorithm also vanishes the chances of starvation. As we have terminated our process only when all the tasks get executed.

The remaining content of this article is organized in different sections. In section II related work has been demonstrates. The next unit is "Problem Formulation" which converses the problem statement. After that, Section IV that is "Proposed Approach: Workflow Scheduling" describes proposed algorithm in detail. Now in Section V "Experimental Results: Proposed Algorithm" portrays the experimental details and imitation outcomes. Lastly, section VI contains "Conclusion and Future scope" which is concluding this article.

II. RELATED WORK

Work process booking issues has been viewed as one of the fundamental difficulties in cloud conditions. Different scientists are chipping away at the booking of work processes on their different compels like restructuring the due date, cost, adjusting the heap on assets and so forth. Numerous calculations were proposed for work process booking and planning the undertakings inside the work processes. A coevolution approach was utilized by Rajkumar buyya [2] to

alter the hybrid approach that is CGA algorithm, which could quicken the meeting and keep the haste. In addition to that, R.Buyya has also projected a versatile penalty work for the striated limitations and other hereditary algorithms. He has utilized an adaptable to address the issue of untimely meeting penalty function without any parameter tuning. In recent years, huge number of scientists has been worked on the same issue. Due to which, some more calculations like Hybrid GA-PSO has planned by Ahmad M. Manasrah [1] to decrease the makespan and the cost and also to adjust the heap of the needy assignments over the heterogonous assets in distributed computing conditions. As per his evaluation, this calculation enhances the heap adjusting of the work process application over the accessible assets. Additionally, Pooja Nagpal [4] exhibits new anticipated calculation which is utilizing advantages of both Enhanced Max-Min and Max-Min calculations together. Test outcomes show that the new proposed calculation speaks to upgraded asset usage with better makespan. One of the examinations of Rajkumar buyya, [3], suggests an asset provisioning and booking technique for logical work processes on Infrastructure as a Service (IaaS) mists. He exhibits a calculation in view of the meta-heuristic enhancement system, molecule swarm advancement (PSO), which plans to limit the general work process execution cost while meeting due date requirements.

III. PROBLEM FORMULATION

In workflows we utilize DAG structure which contains a chain of undertakings. In this pecking order level savvy tasks will be performed. A workflow is representing as a graph, where $\{T_1, T_2, \dots, T_n\}$ are the tasks and presence of edges demonstrates the dependency between two errands. As, shown in figure 1, there is an edge exist amongst T_1 and T_3 which focuses towards T_3 assignments imply that T_1 is a parent and will executed before T_3 .

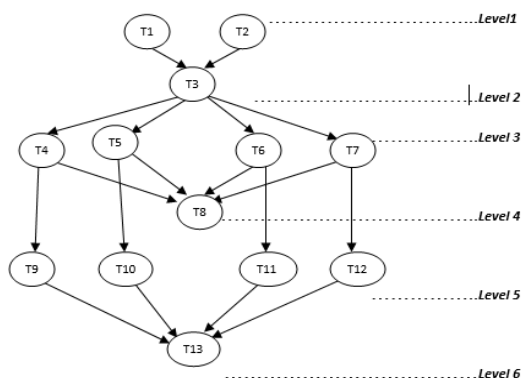


Figure 1: Example of Workflow

Key goal is to plan these assignments though consider that each errand should actualize on scarcely single VM however each virtual machine can perform more than one task. In WorkflowSim, level wise errands are passed to and scheduled that is, after the finishing of first level of undertakings, second level of assignments will be passed for incitement that is, first parents are planned then infant. In our trial we are evaluating

the strength of each assignment based on their makespan. We presume that for each VM kind, the handling limit as far as MIPS is accessible either from the bringer or can be assessed [3]. This data is utilized as a part of our calculation to ascertain the execution time of an endeavor on a given VM.

IV. PROPOSED APPROACH: WORKFLOW SCHEDULING

In our proposed approach, we have performed GA and MINMIN calculations for booking of work processes. With this we have enhance the execution time and makespan of the undertakings allotted on various assets. Here, we have utilized GA for distributing undertakings on various virtual machines and for adjusting the heap on each VM. Next, we have connected MINMIN on each virtual machine which lessened the reaction time of undertakings and afterward, rather than running virtual machines one by one, we have run at that point in parallel, with this holding up time was decreased.

A. Scientific Applications: Workflow Scheduling

Recently, most of the logical applications with their diverse work have process structures. These are represented as DAG structures which are excessively intricate and required enhancement before uploading to the cloud. Montage is a Cloud able galactic and high-vitality material science application that has been utilized to reproject the info pictures, tenacity their experiences, and mosaic them into a solitary picture. The span of a Montage work process relies on the zone of the sky secured by the yield. It is a push to send a convenient, process stern, custom picture mosaicking administration for the space science group [6]. The structure of montage workflow with 25 tasks is shown in Figure 2(a). The CyberShake application is utilized by the Southern California Earthquake Center to portray tremor perils in a locale. These work processes are from the 2011 Production runs which incorporate high recurrence codes. Basic structure of CyberShake workflow has been discussed in Figure 2(b).

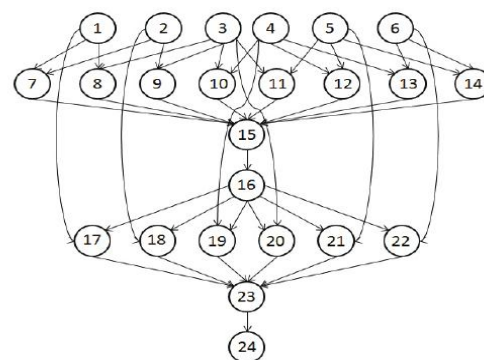


Figure 2(a): Standard Structure of Montage

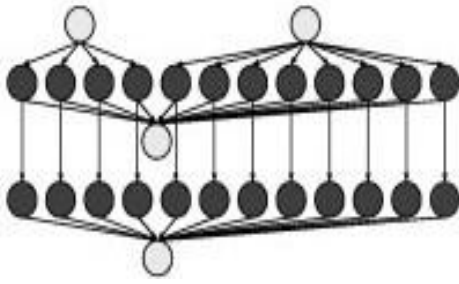


Figure 2(b): Standard Structure of CyberShake

B. General Flow of Proposed Approach:

For the above problem we have proposed a solution, a combination of an evolutionary approach with MINMIN algorithm is used. Genetic Algorithm is a meta-heuristic algorithm, with which we are going to allocate tasks to different virtual machines. After initializing the arbitrary set of solution that is chromosomes, operator's selection, crossover, and mutation are performed on the introduced populace till number of cycles finished.

ALGORITHM 1: GA_MINMIN Approach

Input: Workflow assets for different scientific applications and set of resources
Output: scheduled task
For $i = 0$ to popsize
 $\text{Population}_i \leftarrow \text{randomize}()$
 $\text{Population}_i \leftarrow \text{fitness}()$
End For
While not reach n **do**
 $\text{Chromosome}_j \leftarrow \text{selection}()$
 $\text{Chromosome}_k \leftarrow \text{selection}()$
 $\text{offspring}_p \leftarrow \text{crossover}(\text{Chromosome}_j, \text{Chromosome}_k)$
 $\text{offspring}_{new} \leftarrow \text{mutation}(\text{Offspring}_p)$
 $\text{offspring}_{new} \leftarrow \text{fitness}()$
 If $\text{offspring}_{new} < \text{least fitted chromosome}$
 Swap (offspring_{new} , least fitted chromosome)
 End If
 Repeat
 Select fittest chromosome
 While not reach vmSize **do**
 $\text{VM}_i \leftarrow \text{Sort}(\text{tasks})$
 Repeat
 Run all virtual machines in parallel

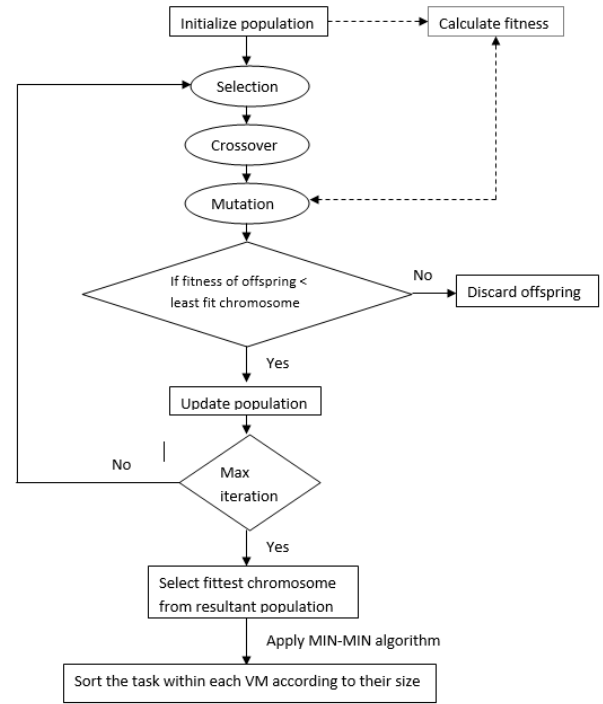


Figure 3: Flow chart of GA and MINMIN algorithm

C. Initialization:

An arbitrarily started populace is passed as a contribution of GA. Populace contains number of chromosomes and every chromosome comprises of number of genes. In work process scheduling, chromosomes are the quantity of errands which we must calendar and traits are the virtual machines.

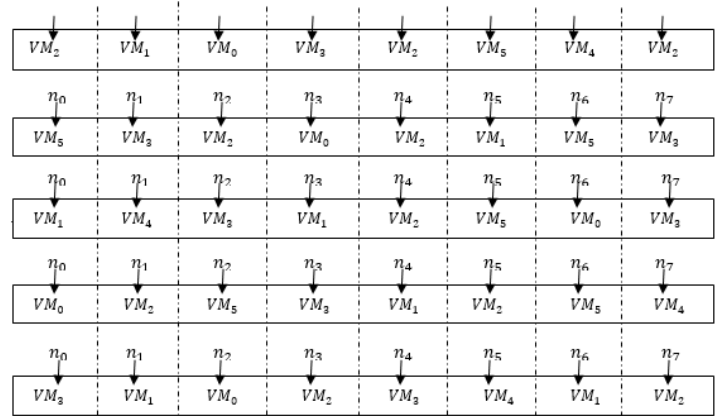


Figure 2: Random population [1]

Figure 2 demonstrates that population has five chromosomes and every chromosome comprises of eight errands and at first which undertakings is assigned to which specific VM. Here, we are taking five virtual machines for instance. Along these lines, eight undertakings are booked on any of the VM at introductory stage. From that point forward, facilitate GA operators are performed on these chromosomes.

D. Fitness Function

The fundamental point of this issue is to upgrade the execution time. So, the size of each undertaking is contemplated and the MIPS value of the VM on which that errands is doled out. The wellness estimation of every chromosome is figured portrayed in equation 1.

$$\sum_{j=1}^k \sum_{i=0}^{n-1} \frac{t_i(\text{size})}{VM_j(\text{mips})} \quad 1$$

n = number of tasks

t = tasks

k = number of virtual machines used

E. Selection

Selection is the initial step which goes for choosing people for multiplication. With a specific end goal to pick the best and the fittest people, choice is done based on wellness of every individual taking part in determination. Determination is the way toward picking two guardians from the populace for intersection. The reason for choice is to underline fitter chromosomes in the populace with the goal that the posterity's henceforth created have higher wellness. For selection, Roulette Wheel Selection (RWS) is used in this we have to select two chromosomes with the help of RWM and perform crossover between them.

F. Crossover

In hybrid activity, two chose chromosomes associate with each other and created an offspring by exchanging their qualities. Here, we are utilizing single point crossover on which the hybrid will happen. An arbitrary number will be produced and utilized as a separation point. As, in crossover genes are exchanged, in workflows tasks exchange their VM's for producing the new individual. New offspring will contain the first half of the one parent and second half of the next parent.

G. Mutation

The yield of the crossover operator has taken as input in mutation. In mutation, genes were swapped within the chromosome, lead to the generation of new offspring. Swapping of qualities implies undertakings changed their virtual machine. For change we have utilized Swap transformation technique for influencing the change to process less composite.

H. Sorting using MINMIN Approach

Check the wellness of out coming about posterity with the instated populace if the wellness is less than the minimum fittest chromosome at that point supplants it with the new offspring. Continue this till the end of iterations and get the fittest population. Now select one best chromosome out of the population and apply MIN-MIN algorithm on it. Now order in which tasks are performed will take into consideration. On each virtual machine sort, the tasks in ascending order of their size. So that task with least size will get the machine first. This will reduce the waiting time of each task.

I. VM Working in Proposed Approach

The key advance that assistance in making our execution much viable was running every virtual machine in parallel. This was done to decrease the response time of undertakings on each VM. As portrayed in Figure 4, at first, a condition has been connected on each VM, that if asset was IDLE, change the status of that VM and make it BUSY and allocate the errand was available in the line in arranged frame. This same procedure will be rehashed till all errands got executed.

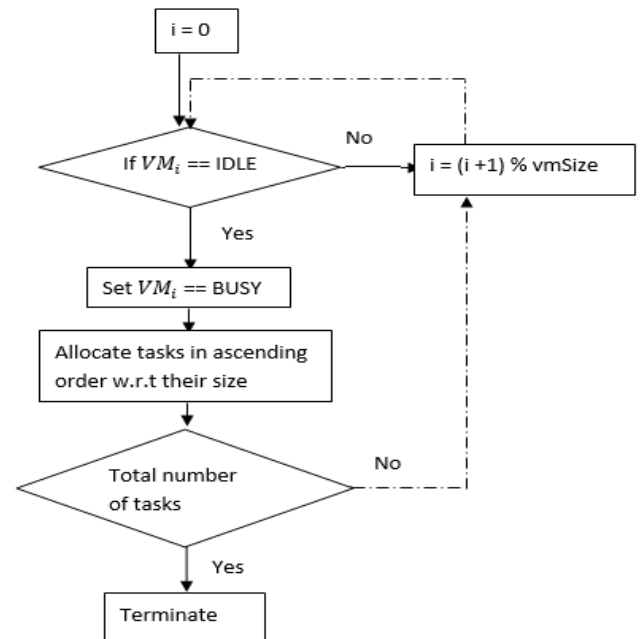


Figure 4: Running VM's in Parallel

V. EXPERIMENTAL RESULTS: PROPOSED APPROACH

Through our test, we were attempting to limit the execution time and reaction time of each assignment in a work process. Essentially, we have dealt with two logical applications, to be specific, Montage and CyberShake. The two applications have different and much complex structures. For assessing logical work processes, we have utilized WorkflowSim to execute these workflows. Here, we have connected our proposed calculation for the planning of youngster hubs by utilizing one of work process scheduler strategy. Additionally, contrasted our proposed calculation and the current methodologies like GA, PSO and Hybrid of GA-PSO.

A. Environment Setup

For assessing our examination, particular parameters have been thought about. As, appeared in Table 1, set of errands between 25 to 1000 have been appointed on 16 virtual machines having distinctive setups for breaking down the proficiency of proposed calculation. The MIPS of each VM differ from 500 to 1500. RAM of every asset has been taken inside the scope of 512 to 2048 MB. Transmission capacity of utilized virtual machines was settled that is 1000. In our proposed algorithm GA algorithm was used for scheduling the

tasks on different virtual machines. GA has been appraised based on different parameters such as different methods used while running different GA operators. As advanced in Table 2, for an instance we have utilized 100 arrangement of populace and connected 100 cycles on it. Two chromosomes were chosen out of 100 populaces with the assistance of Roulette Wheel choice technique. Single point crossover was performed on the selected chromosomes. Then, swap mutation was used to advance the fitness of the chromosome.

Table 1: Prompt Parameters

| Factor | Value |
|--------------------------------|------------|
| Tasks used | 25 – 1000 |
| Number of Virtual Machine used | 16 |
| MIPS | 500 – 1500 |
| RAM | 512 – 2048 |
| Bandwidth | 1000 |

Table 2: GA parameters

| Parameter of GA | Value |
|----------------------|------------------------|
| Size of Population | 100 |
| Number of Iterations | 100 |
| Selection method | Roulette Wheel |
| Crossover | Single point Crossover |
| Mutation | Swap mutation |

B. Proposed Approach on Scientific Applications

Case 1: GA_MINMIN Algorithm on Montage Workflow

On montage work process, execution time was assessed based on specific situations. On the basis the number of tasks, different structures of Montage workflow have been assessed for estimating their execution time. In first situation, we have passed work process structure with less number of assignments that is with 25 undertakings. GA_MINMIN algorithm completed 25 tasks in 82.09 seconds. what's more, in next situations we have are expanding the span of structure by expanding the quantity of assignments with the goal that our proposed calculation will we assessed in a proficient way, publicized in Table 3. Presently, our calculation was assessed with 50 errands and gets executed in 115.97 seconds. Further,

100 and 1000 undertakings were likewise viewed as and get executed in 228.98 and 1580.17 seconds.

Table 3: GA_MINMIN algorithm execution time for Montage workflow

| Workflows tasks | Execution Time |
|-----------------|----------------|
| Montage_25 | 82.09 |
| Montage_50 | 115.97 |
| Montage_100 | 228.98 |
| Montage_1000 | 1580.17 |

A diagram (Figure 5) was plotted utilizing Table 3 which obviously speaks to the pictorial perspective of number of undertakings and their execution time on montage work process.

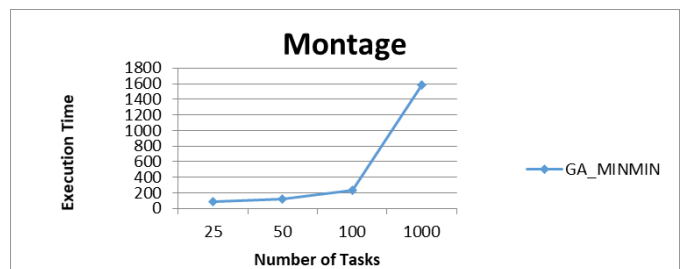


Figure 5: Execution time for Montage workflow using GA_MINMIN algorithm

Case 2: GA_MINMIN Algorithm on CyberShake Workflow

CyberShake is one of the complex and huge structure of logical applications. We were endeavouring to limit the execution time of errands with the assistance of our proposed calculation. GA_MINMIN was connected on various arrangements of assignments of CyberShake work process, as depicted in Table 4. 30 assignments were executed in 275.22 seconds. At that point, we increment the unpredictability of work process by expanding the quantity of undertakings. GA_MINMIN takes 352.98 sec for executing 50 tasks of CyberShake workflow. Moreover, 100 and 1000 errands are additionally assessed and executed inside 583.89 and 2328.76 sec separately. Figure 6 gives the distinctive perspective of execution time of various tasks on CyberShake workflow.

Table 4: GA_MINMIN algorithm execution time for CyberShake workflow

| Workflows tasks | Execution Time |
|-----------------|----------------|
| CyberShake_30 | 275.22 |
| CyberShake_50 | 352.98 |
| CyberShake_100 | 583.89 |
| CyberShake_1000 | 2328.76 |

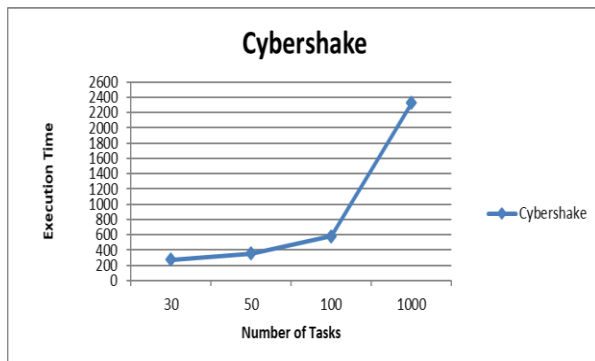


Figure 6: Execution time for CyberShake workflow using GA_MINMIN algorithm

C. Comparison between existing and proposed algorithm

Case 1: GA with GA_MINMIN

Presently, our proposed calculation has been contrasted and existing calculation by taking montage work process in thought. Table 5 demonstrates that GA_MINMIN calculation is more productive then GA as the execution time of our proposed calculation is significantly less than that of GA. In GA algorithm, tasks assigned to each VM were executed in random manner which might increase the waiting time of the task with small duration. But GA_MINMIN algorithm tasks were first sorted in ascending order of their size and the executed. This minimize the waiting time of tasks. For comparing GA with the proposed calculation the execution time of GA is taken from the existing analysis [1].

Table 5: Execution time of GA and GA_MINMIN algorithm

| Algorithm | Number of tasks | Execution time |
|-----------|-----------------|----------------|
| GA_MINMIN | Montage_25 | 82.09 |
| GA | | 197.65 |
| GA_MINMIN | Montage_50 | 115.97 |
| GA | | 250.89 |
| GA_MINMIN | Montage_100 | 228.98 |
| GA | | 345.72 |
| GA_MINMIN | Montage_1000 | 1580.17 |
| GA | | 2402.28 |

Utilizing Table 5 a diagram has been plotted which represent the distinction of execution time between two calculations at various number of assignments. In figure 7 x-hub speaks to the execution time and y-hub speaks to the quantity of undertakings.

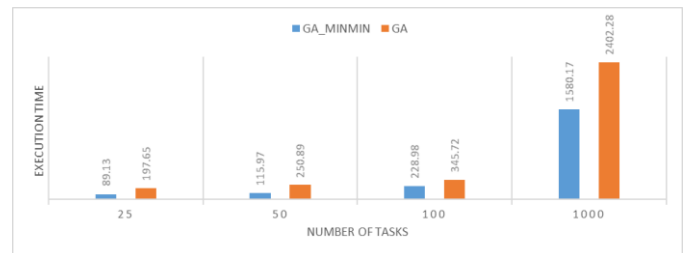


Figure 7: Comparison of GA and GA_MINMIN algorithm based on their execution time.

Case 2: Compare PSO and GA-PSO with GA_MINMIN

GA_MINMIN count was in like manner differentiated and the other transformative estimations like PSO and Hybrid of GA-PSO in view of their execution time. The right examining is depicted in Table 6. The proposed approach gave preferred outcomes over PSO and crossover of GA-PSO calculation proposed in existing analyses [1]. As in GA_MINMIN calculation, usage has been done such that every single virtual machine was running in parallel which diminish the reaction time of the current undertakings. Distinctive situations that diverse number of errands has been utilized for demonstrating our announcement.

Table 6: Execution time of GA_MINMIN, GA-PSO and PSO algorithm

| Algorithm | Number of tasks | Execution time |
|-----------|-----------------|----------------|
| GA_MINMIN | Montage_25 | 82.09 |
| GA-PSO | | 95.09 |
| PSO | | 101.21 |
| GA_MINMIN | Montage_50 | 115.97 |
| GA-PSO | | 116.01 |
| PSO | | 155.31 |
| GA_MINMIN | Montage_100 | 228.98 |
| GA-PSO | | 233.78 |
| PSO | | 253.44 |
| GA_MINMIN | Montage_1000 | 1580.17 |
| GA-PSO | | 1585.06 |
| PSO | | 1802.31 |

As appeared in Table 6 there is less contrast amongst GA_MINMIN and GA-PSO calculation yet at the same time the execution time of GA_MINMIN was less than GA-PSO. In figure 8 we have plotted their comparison.

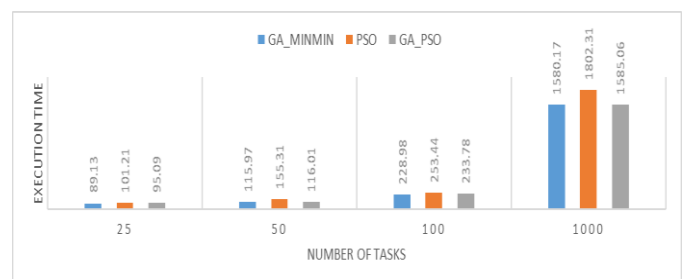


Figure 8: Comparison of GA_MINMIN, GA-PSO and PSO algorithm based on their execution time.

VI. CONCLUSION AND FUTURE WORK:

In this paper, a GA-MINMIN calculation was projected and executed utilizing the WorkflowSim test system for work process assignment planning and for cloud situations. The execution of the proposed calculation was additionally contrasted when referred to these calculations, for example, GA, PSO and GA-PSO. With this we can infer that by actualizing GA and MINMIN calculation in path, it has decreased the response time or execution time of errands on various virtual machines.

In future, we can change the execution stream by moving the undertakings on various VM's at run time. As though any VM is occupied in running a specific errand and another VM is idle without moving around then that task will be moved on the VM which is idle. So the holding up time of the task will be decreased. This will be done at running time. Secondly, in spite of setting irregular portion of assignment to the VM's, we would first be able to check the span of VM and undertaking ought to be apportioned by that.

VII. ACKNOWLEDGEMENT

We acknowledge the Council of Science and Industrial Research (CSIR) Government of India, for funding project (No. 22/693/15/EMR-II) and providing resources to carry out this research work.

VIII. REFERENCES:

1. Ahmad M. Manasrah and Hanan Ba Ali, Workflow Scheduling Using Hybrid GA-PSO Algorithm in Cloud Computing, Published 8 January 2018
2. . Liu, L., Zhang, M., Buyya, R., and Fan, Q. (2016), Deadline- constrained coevolutionary genetic algorithm for scientific workflow scheduling in cloud computing, *Concurrency Computat.: Pract. Exper.*, doi: 10.1002/cpe.3942
3. Maria Alejandra Rodriguez and Rajkumar Buyya, Deadline Based Resource Provisioning and Scheduling Algorithm for Scientific Workflows on Clouds, *IEEE Transactions on Cloud Computing* (Volume: 2, Issue: 2, April-June 1 2014)
4. Poonam Rani, Pooja Nagpal, Optimized Task Scheduling Algorithm for cloud computing environment, *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)* Volume 6, Issue 5, September- October 2017
5. Weiwei Chen and Ewa Deelman, WorkflowSim: A Toolkit for Simulating Scientific Workflows in Distributed Environments, *E-Science (e-Science)*, 2012 IEEE 8th International Conference on 11 January 2013
6. J. C. Jacob, D. S. Katz, T. Prince et al., *The Montage Architecture for Grid-Enabled Science Processing of Large, Distributed Datasets*, Jet Propulsion Laboratory, National Aeronautics and Space Administration, Pasadena, Calif, USA, 2004.
7. H. Magistrale, S. Day, R. W. Clayton, and R. Graves, "The SCEC southern California reference three-dimensional seismic velocity model version 2," *Bulletin of the Seismological Society of America*, vol. 90, no. 6, pp. S65–S76, 2000.
8. Nitish Chopra and Sarbjeet Singh, HEFT based Workflow Scheduling Algorithm for Cost Optimization within Deadline in Hybrid Clouds, *IEEE – 31661* on July 4-6, 2013
9. Mohammad Masdari, Sima ValiKardan, Zahra Shahi, Sonay Imani Azar, Towards workflow scheduling in cloud computing: A comprehensive analysis, *Journal of Network and Computer Applications* 66 (2016) 64–82
10. Harish Garg, A hybrid PSO-GA algorithm for constrained optimization problems, *Elsevier* on 5 December 2015.
11. D. I. George Amalarethinam and T. Lucia Agnes Beena, Differential Evolution Algorithm for Workflow Scheduling (DEWS) in Public Cloud.
12. Anju Bala and Inderveer Chana, Autonomic fault tolerant scheduling approach for scientific workflows in Cloud computing, *Concurrent Engineering: Research and Applications* in 2015
13. Anju Bala and Inderveer Chana, A Survey of Various Workflow Scheduling Algorithms in Cloud Environment, 2nd National Conference on Information and Communication Technology(NCICT) 2011

Speed control of three phase induction motor using neural network

Eng. Nada Mohammed Abd Elmoneim Mahmoud - Faculty of Engineering-Computers Engineering & Control Systems Department- Mansoura University. Demonstrator of Computer at Nile Higher Institute for Commercial Science and Computer Technology. (A.R.E).

E-Mail: Nodafarfosha@gmail.com

Dr. Mohamed Moawad Abdu- Faculty of Engineering-Computers Engineering & Control Systems Department- Mansoura University.

E-Mail: mohmoawad@yahoo.com

Assoc.Prof. Mohammed Sherif Moustafa - Faculty of Engineering-Computers Engineering & Control Systems Department- Mansoura University.

E-Mail: msmksasy@gmail.com

Assoc.Prof. Sabry Fouad Saraya- Faculty of Engineering-Computers Engineering & Control Systems Department- Mansoura University.

E-Mail: drsfsaraya@mans.edu.eg

Abstract

Three phase induction motor Induction is one of the widest spread motor due to its robustness, simple construction, no need for complex circuits for starting. With several available speed control techniques, this paper presents a new Proportional-Integral (PI) controller and Artificial Neural Network (ANNs) control system based on vector control scheme. MATLAB/SIMULINK software may be used to create a 3phase induction engine model. To achieve the effectiveness of the controller, the system is subjected to external disturbance. Experimental results are presented and satisfied with the controller results.

Keywords: Three phases Induction motor, Proportional Integral Controller, Artificial Neural Networks.

1. Introduction

In the last decades, the importance of using the motors has been increased due to their different usage in our daily life or industry [1]. Several motors may be used according to the required performance. Three phase induction motor is one of the widest spread motor due to its robustness, simple construction, no need for complex circuits for starting and suitable for several operated environment [2-4]. Although these advantages, the induction motor suffers from the complexity of controlling the motor speed, low efficiency and power factor at low loads and the nonlinearity of mathematical model [5-6]. Several control system techniques have been developed to control the speed of three phase induction motors, these techniques differ according to efficiency, ease of design, real implementation, reliability and cost [7].

Voltage / Frequency method is the simplest and wide spread way to control the speed of the motor[8-9], but its accuracy is relatively low due to the stator flux

and motor torque are not directly controlled as well as it is normally working without feedback. Vector control method uses several control loops for flux and torque control [10-11], but the highly computational capability is one of the drawbacks of this method [12]. Field control of electrical drives has been one of the famous control techniques for several years [13-15]. In Field-oriented control (FOC) the controlled variables are the stator currents, which transformed to speed dependent system into dq coordinates. These coordinates describe the motor torque and magnetic flux. The control system investigates the reference currents from the motor torque, flux and the desired control speed. This makes the controller accurate and independent of limited bandwidth of the mathematical model [16]. Another control technique is Direct Torque Control (DTC) [17-18]. It is a senseless technique, since the stator flux and motor torque are estimated from the motor voltage and current. The advantages of DTC are faster flux and torque changing responses, a simple control scheme with less calculations and no need to current regulator or co-ordinate transformation. The main drawbacks of DTC are flux ripples and poor dynamic response at low speed and slow transient response at starting [19].

Different implementation algorithms and controllers were developed to control the speed of induction motor as PID control [20-22], Pulse width modulation (PWM)[23-24], fuzzy logic[25-27], Artificial Neural Network[28-30], sliding mode control[31-33] and Particle swarm optimization (PSO)[34-36].

This paper presents a new PI-Artificial Neural Network (PI-ANN) controller technique for speed tracking of three phase induction motor based on vector control scheme. To validate the effectiveness of the controller, the motor is subjected to multi disturbances as torque disturbances and speed variations during operation.

2. Mathematical model of three phase induction motor

The stator-to-rotor coupling terms are a function of the rotor position, so when the rotor rotates, the coupling terms changing with time. To solve this problem, induction motor equations are transferred to the quadrature rotating reference frame such that the mutual inductances are not time dependent. The model equations are shown from equation 1 to 6 [37]. The equivalent circuit diagrams for each phase of an IM is shown in Fig. 1

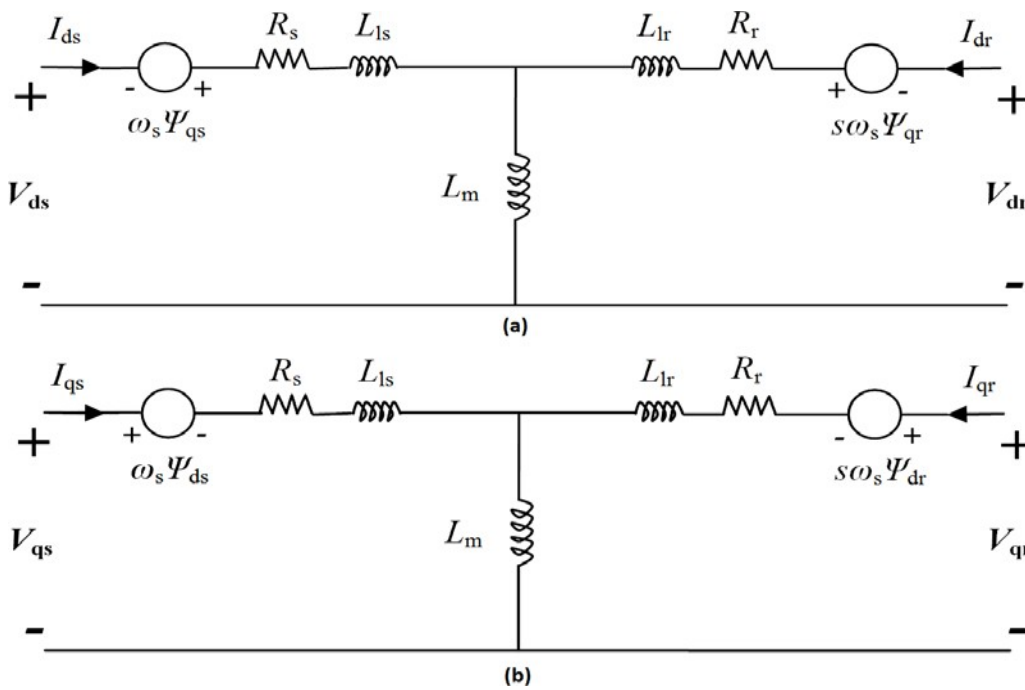


Fig.1: Induction Motor Circuit Diagram

Stator Model Equations:

$$V_{qs} = R_s * i_{qs} + \frac{d\varphi_{qs}}{dt} + \omega_e * \varphi_{ds} \quad (1)$$

$$V_{ds} = R_s * i_{ds} + \frac{d\varphi_{ds}}{dt} + \omega_e * \varphi_{qs} \quad (2)$$

Rotor Model Equations:

$$V_{qr} = R_r * i_{qr} + \frac{d\varphi_{qr}}{dt} + (\omega_e - \omega_r) * \varphi_{dr} \quad (3)$$

$$V_{dr} = R_r * i_{dr} + \frac{d\varphi_{dr}}{dt} + (\omega_e - \omega_r) * \varphi_{qr} \quad (4)$$

Where:

V_{qs}, V_{ds} : quadrature and direct axes stator voltages.

V_{qr}, V_{dr} : quadrature and direct axes rotor,

R_s, R_r : stator and rotor resistance.

i_{qs}, i_{ds} : quadrature and direct axes stator currents.

i_{qr}, i_{dr} : quadrature and direct axes rotor currents.

$\varphi_{qs}, \varphi_{ds}$: quadrature and direct axes stator flux.

$\varphi_{qr}, \varphi_{dr}$: quadrature and direct axes rotor flux.

ω_e : electrical rotor angular velocity (rad/sec).

ω_r : rotor speed (rad/sec).

The development torque by interaction of air gap flux and rotor current can be found as:

$$T_e = \frac{3}{2} * \frac{P}{2} * \vec{\varphi_m} * \vec{I_r} \quad (5)$$

By resolving the variables into d-q components:

$$T_e = \frac{3}{2} * \frac{P}{2} * (\varphi_{ds} * i_{qs} - \varphi_{qs} * i_{ds}) \quad (6)$$

Figure 2 shows the MATLAB Simulink induction motor model based indirect vector control.

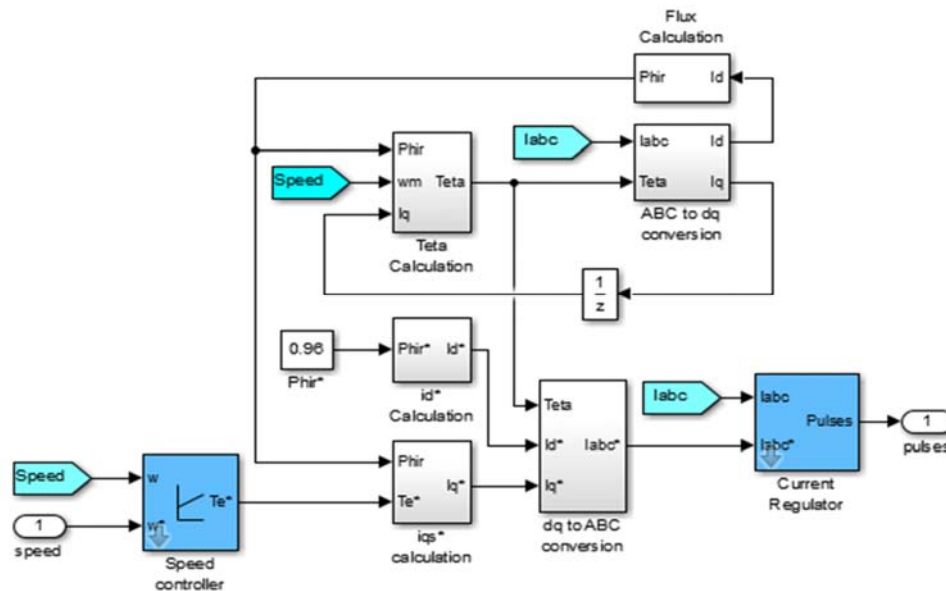


Fig.2: Indirect Vector Controller Developed of Three Phase Induction Motor

Because of the nonlinearity and tracking speed of the three phase induction motor, it is required to design a new controller that can overcome these problems.

3. Artificial Neural Network

ANN is computational model that attempts to mimic the construction and tasks of neural networks and specify simple approximations to parts of real brains [38-39] Requisite structure block of ANN is neuron, it is a simple function. A model consists of three groups of principles as shown Fig.3: multiplication, summation and activation. ANN consists of elements connected together to perform a specific task. At the beginning of ANN each input is multiplied with weight. In the second stage there is a summation function that collects all weighted inputs and bias. At the end, all weighted inputs and bias are exceeding through transfer function. Activation function

used for converting the activation level to obtain the output of neurons. Note that each input may be external or the output of some other neuron

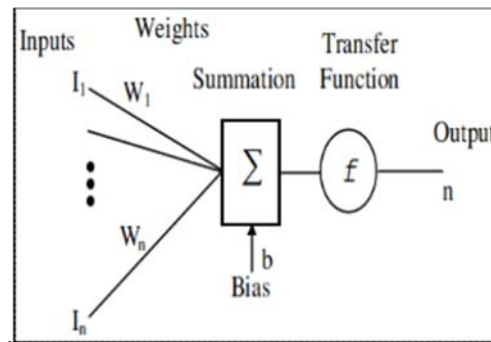


Fig.3: Working principle of ANN

The most important type of ANNs used widely today are multilayer perceptrons. Multilayer perceptrons are class of constructions called feed forward neural networks as shown in Fig. 4. MLP have been used in microwave and optimization. In MLP, the neurons are classified in to layers. In MLP construction, the first layers called input layers and the last layers called output layers. Input and output layers of MLP perform the all network.

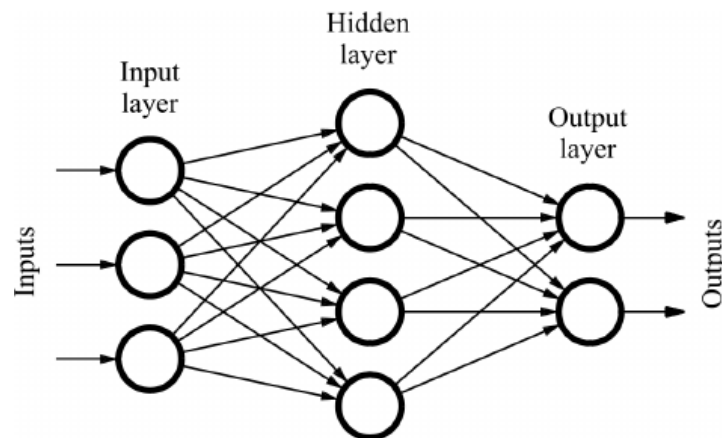


Fig.4: Multilayer perceptron feed forward construction

The number of neurons needed for modeling problem still an open question. There is not obvious answer, the number of neurons established on the degree of nonlinearity and the dimensionality of the task. Nonlinear tasks need more neurons and softer tasks need fewer neurons.

Artificial neural network has several advantages as: it deals with the system don't care to the nonlinearity in the model, rapidly implementation speed, Adapted with environment. But it has some drawbacks as: needs long time for training, too hard to perform problems that relate impaction of paradigm and memory.

4. The Suggested PI-ANN Controller

A new PI-ANN controller is designed for speed tracking of IM. The Matlab Simulink model including the controller is shown in Fig.5. The electrical parameters

are summarized in table 1. The controller depends on use two control signals to investigate the reference motor angular position and hence the required speed. The first control signal is produced from PI controller, which receives the reference required speed and the actual motor speed for fast speed tracking. The second control signal comes from the ANN. The Advantages of the proposed neural network controller that it deals with the IM regardless the model nonlinearity, it has parallel processing and generalization capacities. In Simulink library PWM inverter is built using a Universal Bridge block, connected with a 780 volts dc source, the motor torque 300 Nm. Mechanical load of motor drives is characterized by the inertia, friction coefficient, and load torque.

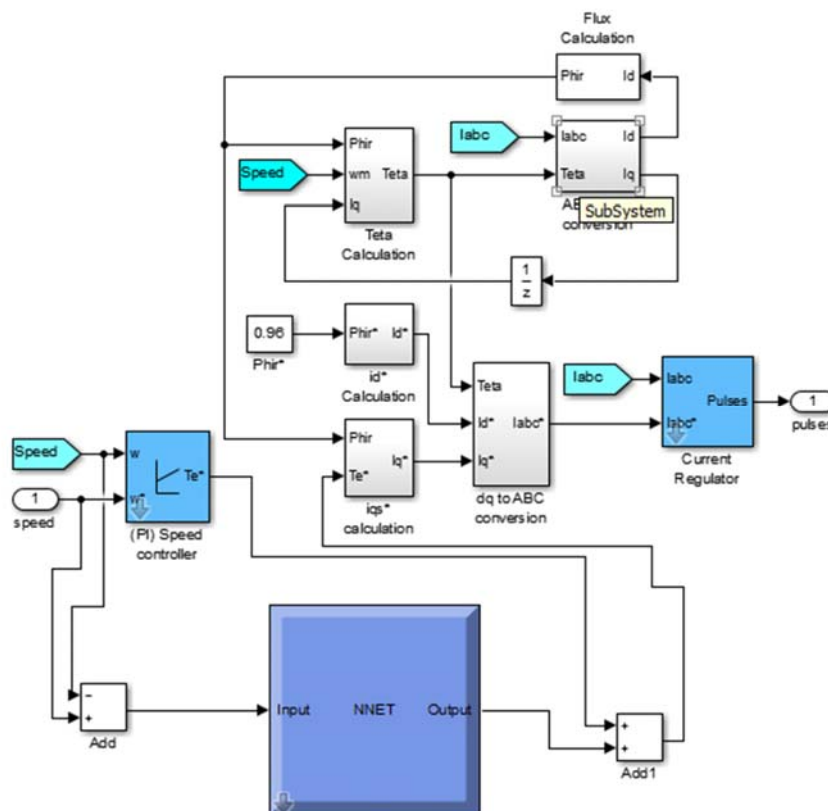


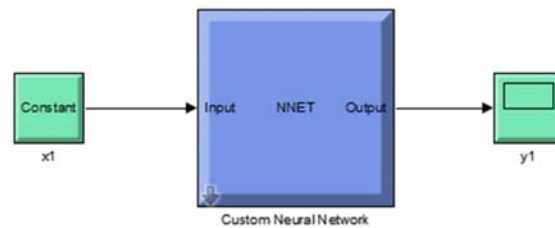
Fig 5. Three Phase Induction Motor Controller Using PI-ANN.

Table 1: Induction Machine Electrical Parameters

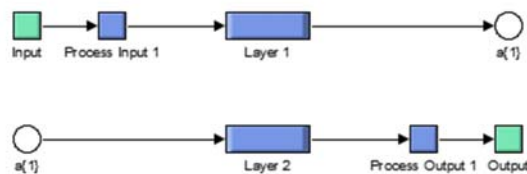
| | |
|---|----------|
| P (Power) | 50 HP |
| V (voltage) | 460 V |
| S (Speed of motor) | 1750 RPM |
| R _s (the Resistance of stator) | 0.087Ω |
| L _s (the inductance of stator) | 0.8 mH |
| R _r (the resistance of rotor) | 0.228 Ω |
| L _r (the inductance of rotor) | 0.8 mH |
| the Inertia (J) | 1.662 |
| friction coefficient (B) | 0.1 |
| No of poles | 2 |

Construction of neural network controller selected is shown in Fig.6. The controller input is the speed error signal (e). The Input and hidden layer activation functions are logarithmic sigmoid and linear for the output layer.

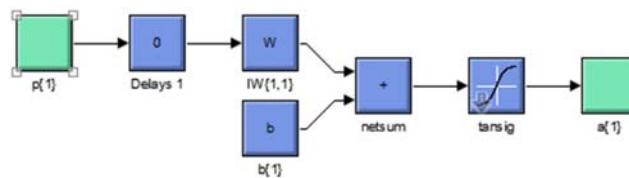
Now the Matlab simulation model of developed NN controller .



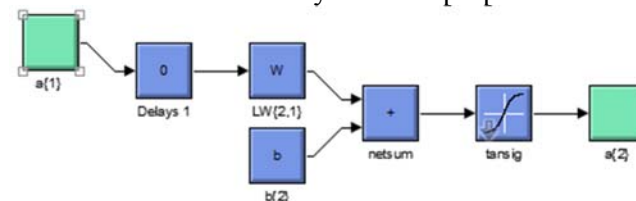
(a) Simulation model of developed NN controller



(b) The internal structure of proposed NN controller.



c) The internal structure of layer one of proposed NN controller.



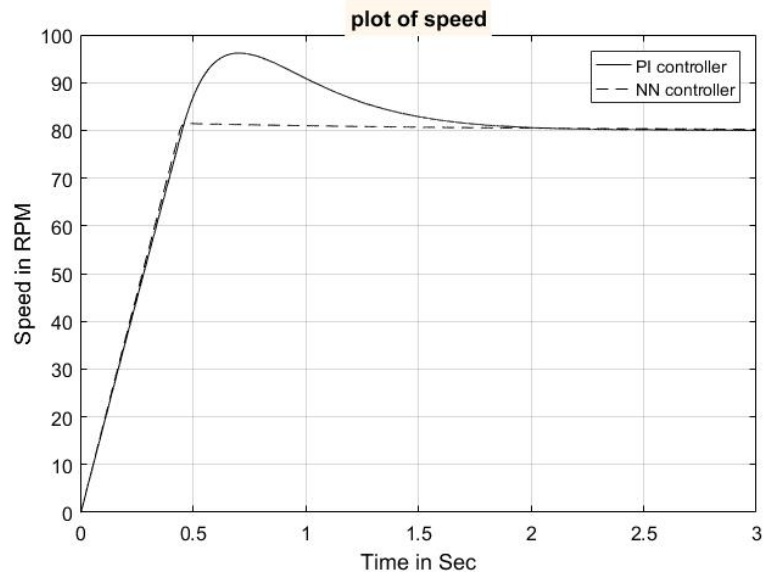
D) Internal structure of the second layer of proposed NN controller.

Fig 6. Simulation model of developed ANN controller

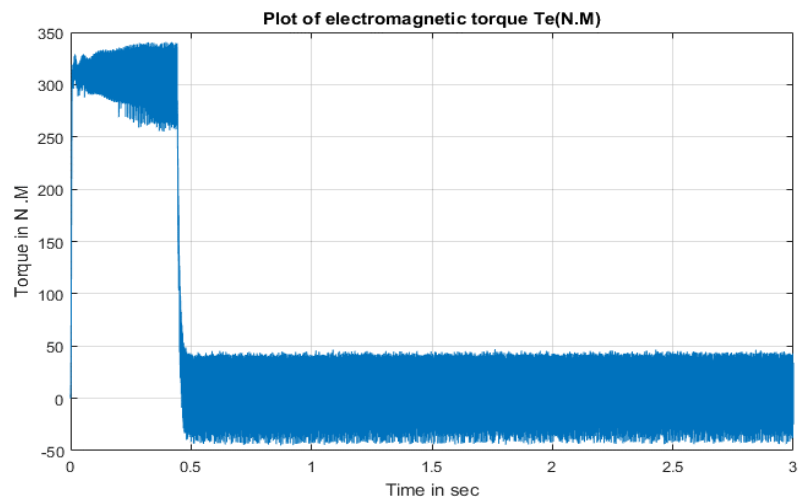
5. Simulation Results

This section shows the controller speed tracking for several reference speeds.

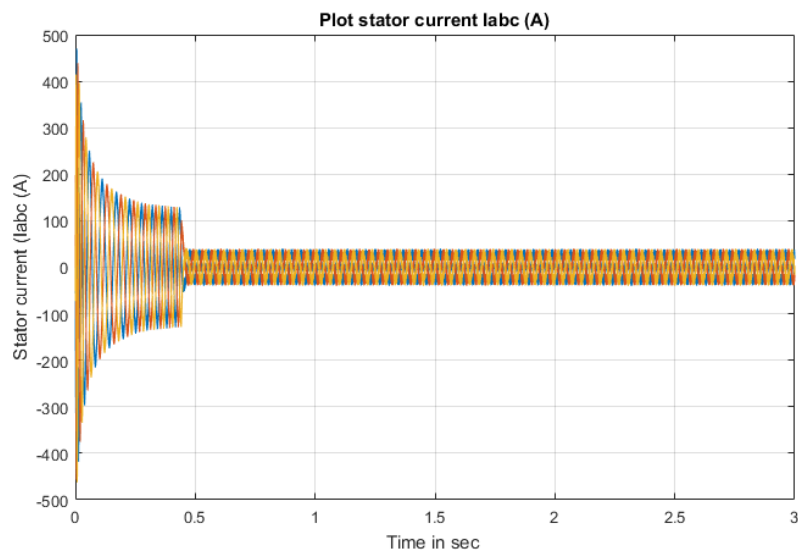
Case 1: the reference speed is 80 RPM, the motor speed, torque and currents are shown in Fig.7.



(a)



(b)

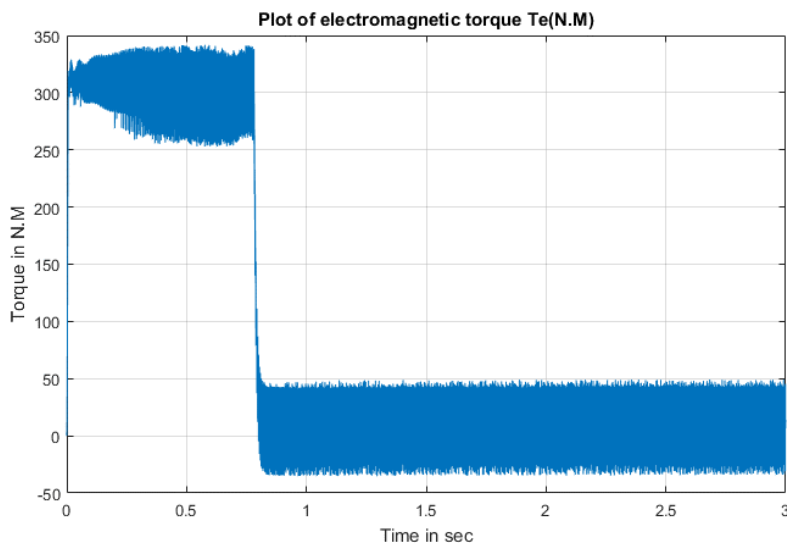
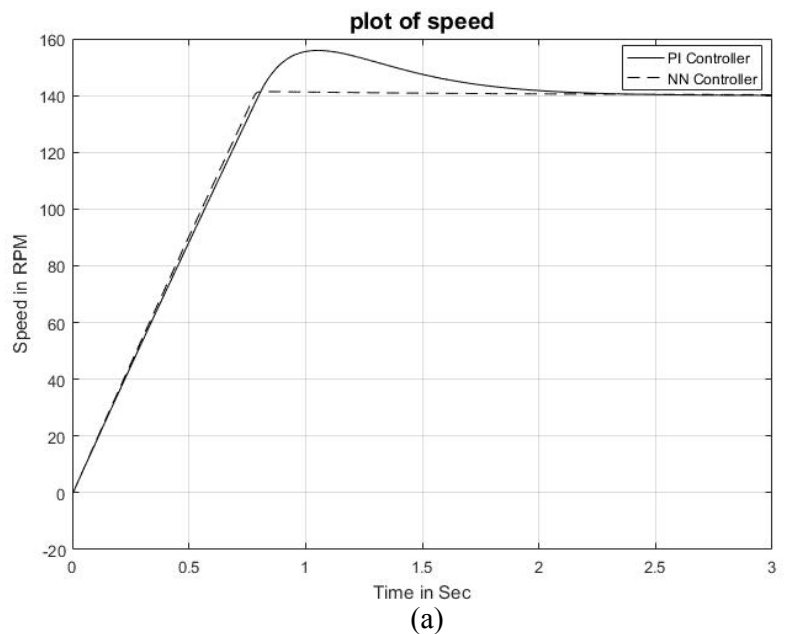


(c)

Fig 7: The performance of IM Current for Reference Speed = 80 for (speed, torque and currents)

Fig.5 shows that, the PI-ANN controller has better performance compared with PID controller. The speed overshoot has been decreased and the reaching time is 1.4 sec. While in PID controller, the reaching time is 2.2 sec with 18 RPM overshooting speed.

Case2: For Reference Speed = 140 RPM. The motor speed, torque and currents are shown in Fig.8.



(b)

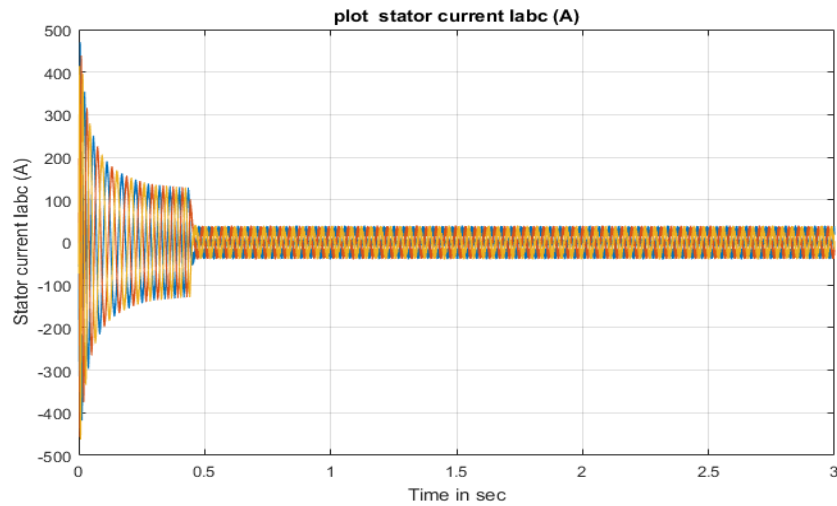


Fig 8: The performance of IM Current for Reference Speed = 140 for (speed, torque and currents)

As shown in Fig.8, the motor reaches the reference speed using PI-ANN controller faster than using PI controller with less speed overshooting. Table 2 summarizes the results op PI and PI-ANN controllers.

Table 2: Comparison PI controller and NN controller performance specifications.

| PI Controller | Reference Speed 80 RPM | Reference Speed 140 RPM |
|-------------------|---------------------------|----------------------------|
| Settling Time | 2.2 sec | 2.3 sec |
| Rise Time | 0.3sec | 0.6 sec |
| Peak Time | 0.7sec | 1 sec |
| Maximum overshoot | 18 RPM | 18 RPM |
| PI-ANN Controller | Reference Speed 80 RPM | Reference Speed 140 RPM |
| Settling Time | 1.4 sec | 1.6 sec |
| Rise Time | 0.2 sec | 0.5 sec |
| Peak Time | 0.4 sec | 0.8 sec |
| Maximum overshoot | 1 RPM | 0.4 RPM |

When we use PI controller for reference speed=80RPM the system suffers from high overshoot with about 18RPM and take 2.2sec to stabilize with rising time 0.3sec. For reference speed=140RPM the system suffers from about 18RPM overshoot and take 2.3sec to stabilize with rising time=0.6sec.

While when we use PI-ANN controller for reference speed=80RPM the overshoot of the system will be 1RPM and take 1.4sec to stabilize with rising time=0.2sec. For reference speed=140RPM the overshoot of the system will be 0.4RPM and take 1.6sec to stabilize with rising time=0.5sec.

For achieving the effectiveness of the controller, the motor will be subjected to external disturbances as:

Case 3: Changing the reference speed during the operation, it changes from 80RPM to 120RPM then to 100RPM. Figure 9 shows the disturbed IM model. Figure 10, shows the motor curves at this disturbance.

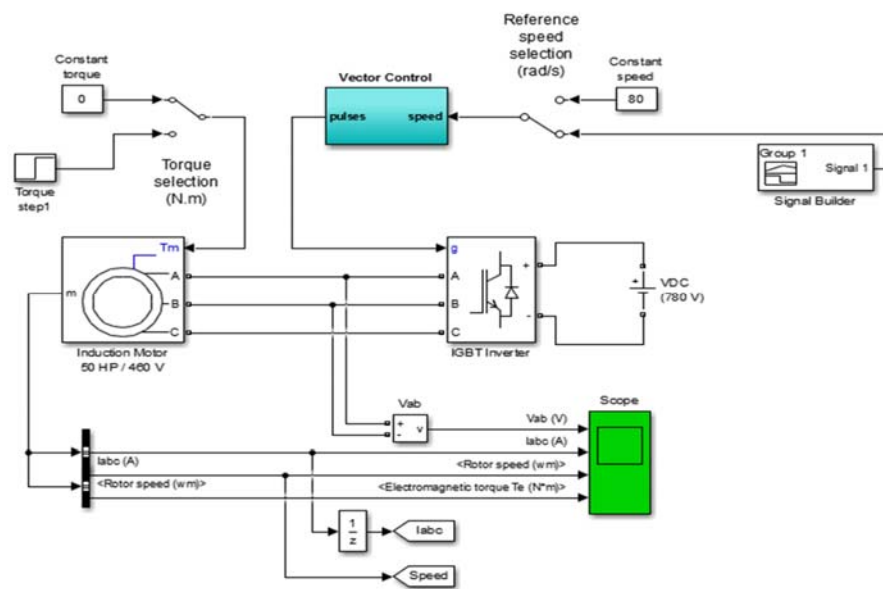
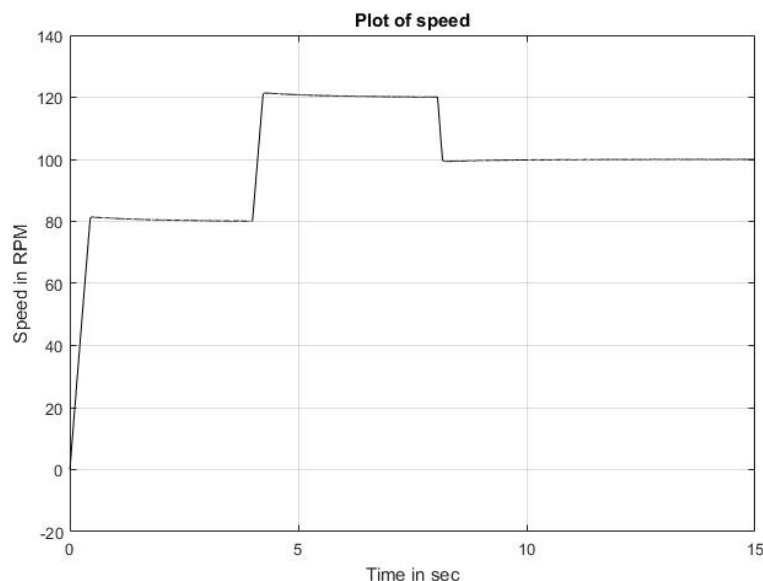


Fig 9: The Disturbed IM model using PI-ANN controller.



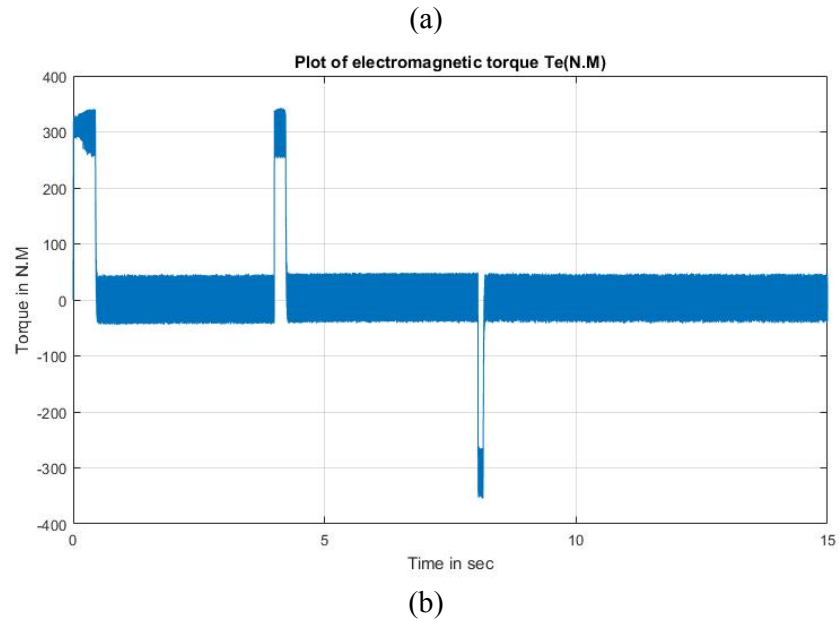
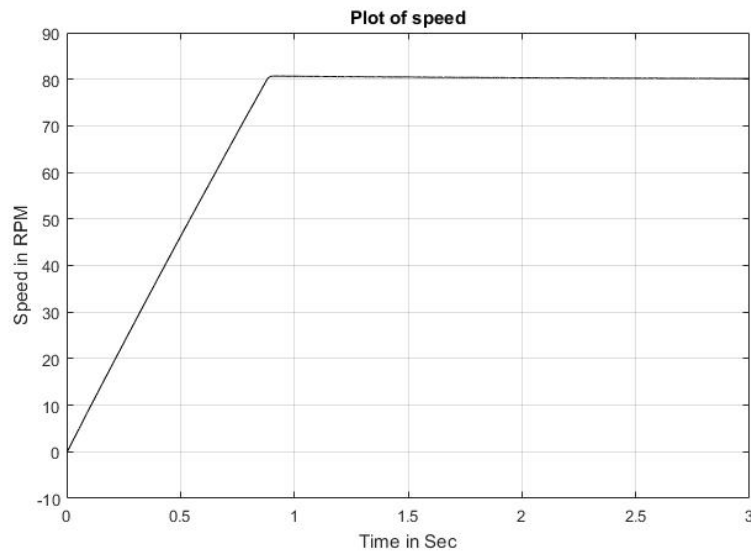


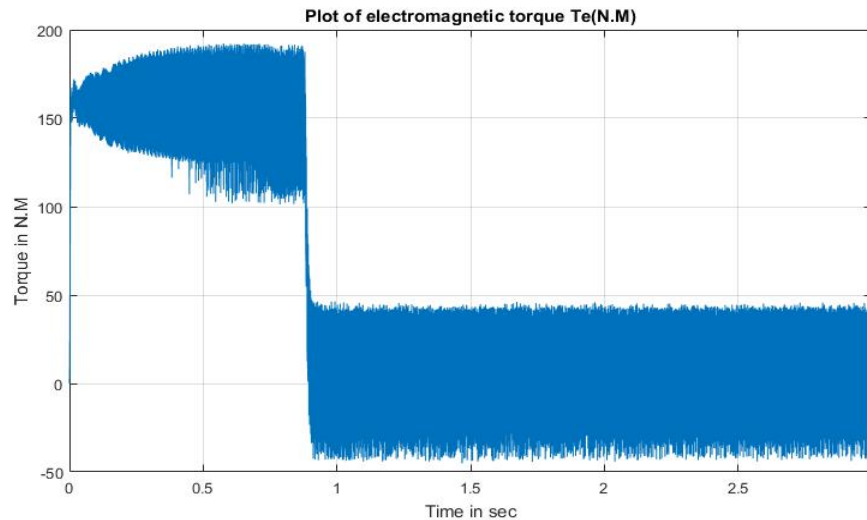
Fig.10: The disturbed reference speed controller results (Speed, Torque)

The motor takes 0.6sec to reach to 80RPM then it takes 0.3sec to reach from 80 RPM to 120 RPM. The system takes 0.2 sec to stabilize without overshoot.

Case 4: Other disturbance, the motor load torque is changed to 150N.M at reference speed 80RPM. Figure 11, shows the motor curves at the torque disturbance.



(a)



(b)

Fig.11: The disturbed load torque controller results (Speed, Torque)

When we make torque disturbance =150N.M, the motor stabilizes at 80 RPM after 1 sec. with no overshooting and not affected by torque disturbance.

6. Conclusion.

Several control techniques are used to control the speed of three phase induction motors. Vector control Technology is a quite control method on the powerful induction motor velocity control system. This paper presents a new controller design using PI-ANN design with Vector control Technology. The controller was tested at different speed signals with less overshooting and fast response than the classical PI controller. The effectiveness of the controller was achieved by expose the motor to external disturbances as changing the reference speed signal and reducing the load torque. The controller was designed by Matlab Simulink. The experimental results are presented and agreed with the simulation results.

7. References

1. A. Hughes, "Electric Motors and Drives Fundamentals, Types and Applications", Third edition 2006, Published by Elsevier Ltd., ISBN-13: 978-0-7506-4718-2, 2006.
2. I. Takahashi and T. Noguchi "A new quick-response and high efficiency control strategy of an induction motor", IEEE Trans Ind. Appl., Vol.22 (5), pp. 820-827, 1986.
3. O. Barambones and P. Alkorta, "A robust vector control for induction motor drives with an adaptive sliding-mode control law", Journal of the Franklin Institute, Vol. 348 (2), Pages 300-314, 2011.
4. O. Barambones, A. Garrido and I. Garrido, "Robust speed estimation and control of an induction motor drive based on artificial neural networks", International Journal of Adaptive Control and Signal Processing, Vol. 22 (5), P.p: 440-464, 2008.
5. M. Ghanes , J. Barbot , J. De Leon and A. Glumineau, "A robust sensorless output feedback controller of the induction motor drives: new design and experimental

- validation”, Journal International Journal of Control, Vol. 83 (10), P.p: 484-497, 2010.
6. S. Jeon, D. Baang, and J. Choi, “Adaptive Feedback Linearization Control Based on Airgap Flux Model for Induction Motors”, International Journal of Control, Automation, and Systems, Vol. 4 (4), pp.414-427, 2006.
 7. G. Ramírez, L. Valdés, M. Medina, C. Beltrán, and C. López, “Adaptive nonlinear control of induction motor”, International Journal of Control, Automation and Systems, Vol. 9 (1), p.p: 176–186, 2011.
 8. S. Kailaswar and R. Keswani, “Speed Control of Three Phase Induction Motor by V/f Method for Batching Motion System”, International Journal of Engineering Research and Applications, Vol. 3, p.p.1732-1736, 2013.
 9. V. Yendole and P. R.Jawale, “Speed Control of Three Phase Induction Motor by Using V/F Method”, International Journal of Research in Advent Technology, National Conference “CONVERGENCE 2016”, p.p. 115-119, 2016.
 10. O. Barambones and P. Alkorta, “Vector control for induction motor drives based on adaptive variable structure control algorithm”, Asian Journal of Control, Vol.12 (5), P.p: 640–649, 2010.
 11. P. Kjaer, T. Kjellqvst and C. Delaloye, “Estimation of field current in Vector controlled Synchronous Machine Variable Speed Drives Employing Brushless Asynchronous Exciters”, IEEE Trans. on Industry Applications, Vol. 41(3), P.p: 834–840, 2005.
 12. A. Chourasia, V. Srivastava, A. Choudhary and S. Praliya, “Comparison study of Vector Control of Induction Motor Using Rotor Flux Estimation by Two Different Methods”, International Journal of Electronic and Electrical Engineering, Vol.7 (3),P.p: 201-206, 2014
 13. S. Sathikumar, J. Vithayathil, “Digital simulation of field-oriented control of induction motor”. IEEE Trans. Ind. Electron. Vol. 31, P.p:141–148, 1984
 14. A. Kumar and T. Ramesh, “Direct Field Oriented Control of Induction Motor Drive”, Advances in Computing and Communication Engineering (ICACCE), 2015 Second International Conference on Advances in Computing and Communication Engineering , 1-2 May 2015
 15. R.D.Lorendz, “Tuning of field oriented induction motor controller for high performance applications”, IEEE transactions Ind. Appl.Vol. 31 no. 4,pp. 812-822, 1995.
 16. J. Hotz and E. Bube, “Field oriented asynchronous PWM for high performance AC machine drives operating at low switching frequency” IEEE Trans. on Industry Applications, Vol. 27(3), P.p:574–581, 1991.
 17. J. Rodríguez, J. Pontt, S. Kouro and Pablo Correa “Direct torque control with imposed switching frequency in an 11-level cascaded inverter” IEEE Trans Ind. Electron, Vol. 51 (4), pp. 827-833, 2004.
 18. A. Zemmit, S. Messalti, and A. Harrag, “A new improved DTC of doubly fed induction machine using GA-based PI controller”, Ain Shams Engineering Journal , Available online 20 February 2017.
 19. H. Liu and H. Zhang, “A novel direct torque control method for brushless DC motors based on duty ratio control”, Journal of the Franklin Institute, Vol. 354 (10), P.p. 4055-4072, 2017.

20. D. Asija, "Speed control of induction motor using fuzzy-PI controller", 2010 2nd International Conference on Mechanical and Electronics Engineering, 1-3 Aug, 2010.
21. S. Senthilkumar and S. Vijayan, "Simulation of High Performance PID Controller for Induction Motor Speed Control with Mathematical Modeling", Research Journal of Applied Sciences, Engineering and Technology, Vol. 6 (18), P.p:3343-3348, 2013.
22. A. Aggarwal, J. N. Rai and M. Kandpal, "Comparative Study of Speed Control of Induction Motor Using PI and Fuzzy Logic Controller", Journal of Electrical and Electronics Engineering, Vol. 10 (2) , P.p: 43-52, 2015.
23. P. S. Joshi and A. M. Jain, "Wireless Speed Control Of An Induction Motor Using PWM Technique With GSM", International Journal of Engineering Research & Technology (IJERT) ,Vol. 2 (5), P.p: 2278-0181, 2013.
24. P. Wei, j. Yu, F. Shi, X. Wei, Y. Wang and Q. Zhao, "The PWM Control of the Three-phase Induction Motor", International Conference on Manufacturing Science and Engineering (ICMSE 2015), Atlantis Press, P.p: 842-845, 2015.
25. S. Yuan, C. Tseng and C. Chiu, "Design of a novel adaptive TSK-fuzzy speed controller for use in direct torque control induction motor drives", Applied Soft Computing, Vol. 31, P.p: 396-404, 2015.
26. F. Areed, A. Haikal, and R. Mohammed, "Adaptive neuro-fuzzy control of an induction motor", Ain Shams Engineering Journal, Vol. 1 (1), P.p:71-78, 2010.
27. F. Rashidi, "Sensorless speed control of induction motor drives using robust and adaptive neuro-fuzzy based intelligent controller", IEEE international conference on industrial technology (ICIT), 2004, P.p. 617-627, 2004.
28. M.Congli1, H. Wentao, Y. Kaiting and L. Guohai , "Speed-Regulating System for Induction Motor and Inverter based on Hammerstein Model and Neural Network Control", International Journal of Control and Automation, Vol. 8 (3), P.p:271-286, 2015.
29. G.Rajamany and S. Srinivasan, "An Artificial Neural Networks Application for the Automatic Detection of Severity of Stator Inter Coil Fault in Three Phase Induction Motor", Journal of Electrical Engineering and Technology, Vol. 12 (6), P.p: 2219-2226, 2017.
30. V. Jamuna and S. Reddy, "Artificial Neural Network based Speed Control of Bidirectional Chopper fed Induction Motor Drive using DFRT Theory", Asian Power Electronics Journal, Vol. 4 (1) , P.p: 36-42, 2010.
31. S. Rashinkar and S. Sankeshwari, "Sliding-Mode Controller Based Indirect Vector Control Of Induction Motor", International Journal of Advances in Engineering & Technology, Vol. 7 (5), P.p.: 1470-1475, 2014.
32. A. Kumar , V. Tripathi and A. Tiwari, "Sliding Mode Vector Control Of Three Phase Induction Motor", International Journal of Engineering Research & Technology, Vol. 2 (5), P.p: 1910-1914, 2013.
33. J.B.Oliveiraa, A.D.Araujob and S.M.Dias, "Controlling the speed of a three-phase induction motor using a simplified indirect adaptive sliding mode scheme", Control Engineering Practice, Vol. 18 (6), P.p: 577-584, 2010.
34. M. M. Abd-Elsalam, M. S. Elkasas, S. F. Saraya and A. H. Awad," Enhancing Tracking Performance Parameters of Induction Motor Based on PI-PSO Algorithm", International Journal of Scientific & Engineering Research, Volume 8, Issue 5,pp. 1685-1702, 2017.

35. S. Mahapatra, R. Daniel, D. Narayan and S. Nayak, "Induction Motor Control Using PSO-ANFIS", International Conference on Intelligent Computing, Communication & Convergence (ICCC-2014), procedia computer science, P.p: 753-768, 2015.
36. M.M. Eissa , G. S. Virk , A. M. AbdelGhany , E. S. Ghith, "Optimum Induction Motor Speed Control Technique Using Particle Swarm Optimization", International Journal of Energy Engineering , Vol. 3(2), P.p: 65-73, 2013.
37. P. L. Ratnani, A. G. Thosar, "Mathematical Modeling of a 3 Phase Induction Motor Using MATLAB/Simulink", International Journal of Modern Engineering Research, Vol. 4 (6), P.p: 62-67, 2014.
38. S.Wang,W. Chaovalitwongse and R. Babuska," Machine Learning Algorithms in Bipedal Robot Control", IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)", Vol. 42, P.p:728-743, 2012.
39. R. H. Nielsen. "Theory of Backpropagation Neural Network", International Joint Conf. on Neural Networks, P.p: 1585-1592, 1989.

IJCSIS REVIEWERS' LIST

Assist Prof (Dr.) M. Emre Celebi, Louisiana State University in Shreveport, USA
Dr. Lam Hong Lee, Universiti Tunku Abdul Rahman, Malaysia
Dr. Shimon K. Modi, Director of Research BSPA Labs, Purdue University, USA
Dr. Jianguo Ding, Norwegian University of Science and Technology (NTNU), Norway
Assoc. Prof. N. Jaisankar, VIT University, Vellore, Tamilnadu, India
Dr. Amogh Kavimandan, The Mathworks Inc., USA
Dr. Ramasamy Mariappan, Vinayaka Missions University, India
Dr. Yong Li, School of Electronic and Information Engineering, Beijing Jiaotong University, P.R. China
Assist. Prof. Sugam Sharma, NIET, India / Iowa State University, USA
Dr. Jorge A. Ruiz-Vanoye, Universidad Autónoma del Estado de Morelos, Mexico
Dr. Neeraj Kumar, SMVD University, Katra (J&K), India
Dr. Genge Bela, "Petru Maior" University of Targu Mures, Romania
Dr. Junjie Peng, Shanghai University, P. R. China
Dr. Ilhem LENGILIZ, HANA Group - CRISTAL Laboratory, Tunisia
Prof. Dr. Durgesh Kumar Mishra, Acropolis Institute of Technology and Research, Indore, MP, India
Dr. Jorge L. Hernández-Ardieta, University Carlos III of Madrid, Spain
Prof. Dr. C. Suresh Gnana Dhas, Anna University, India
Dr. Li Fang, Nanyang Technological University, Singapore
Prof. Pijush Biswas, RCC Institute of Information Technology, India
Dr. Siddhivinayak Kulkarni, University of Ballarat, Ballarat, Victoria, Australia
Dr. A. Arul Lawrence, Royal College of Engineering & Technology, India
Dr. Wongyos Keardsri, Chulalongkorn University, Bangkok, Thailand
Dr. Somesh Kumar Dewangan, CSVTU Bhilai (C.G.) / Dimat Raipur, India
Dr. Hayder N. Jasem, University Putra Malaysia, Malaysia
Dr. A.V. Senthil Kumar, C. M. S. College of Science and Commerce, India
Dr. R. S. Karthik, C. M. S. College of Science and Commerce, India
Dr. P. Vasant, University Technology Petronas, Malaysia
Dr. Wong Kok Seng, Soongsil University, Seoul, South Korea
Dr. Praveen Ranjan Srivastava, BITS PILANI, India
Dr. Kong Sang Kelvin, Leong, The Hong Kong Polytechnic University, Hong Kong
Dr. Mohd Nazri Ismail, Universiti Kuala Lumpur, Malaysia
Dr. Rami J. Matarneh, Al-isra Private University, Amman, Jordan
Dr. Ojesanmi Olusegun Ayodeji, Ajayi Crowther University, Oyo, Nigeria
Dr. Riktesh Srivastava, Skyline University, UAE
Dr. Oras F. Baker, UCSI University - Kuala Lumpur, Malaysia
Dr. Ahmed S. Ghiduk, Faculty of Science, Beni-Suef University, Egypt
and Department of Computer science, Taif University, Saudi Arabia
Dr. Tirthankar Gayen, IIT Kharagpur, India
Dr. Huei-Ru Tseng, National Chiao Tung University, Taiwan
Prof. Ning Xu, Wuhan University of Technology, China
Dr. Mohammed Salem Binwahlan, Hadhramout University of Science and Technology, Yemen
& Universiti Teknologi Malaysia, Malaysia.
Dr. Aruna Ranganath, Bhoj Reddy Engineering College for Women, India
Dr. Hafeezullah Amin, Institute of Information Technology, KUST, Kohat, Pakistan

Prof. Syed S. Rizvi, University of Bridgeport, USA
Dr. Shahbaz Pervez Chattha, University of Engineering and Technology Taxila, Pakistan
Dr. Shishir Kumar, Jaypee University of Information Technology, Wakanaghat (HP), India
Dr. Shahid Mumtaz, Portugal Telecommunication, Instituto de Telecomunicações (IT) , Aveiro, Portugal
Dr. Rajesh K Shukla, Corporate Institute of Science & Technology Bhopal M P
Dr. Poonam Garg, Institute of Management Technology, India
Dr. S. Mehta, Inha University, Korea
Dr. Dilip Kumar S.M, Bangalore University, Bangalore
Prof. Malik Sikander Hayat Khiyal, Fatima Jinnah Women University, Rawalpindi, Pakistan
Dr. Virendra Gomase , Department of Bioinformatics, Padmashree Dr. D.Y. Patil University
Dr. Irraivan Elamvazuthi, University Technology PETRONAS, Malaysia
Dr. Saqib Saeed, University of Siegen, Germany
Dr. Pavan Kumar Gorakavi, IPMA-USA [YC]
Dr. Ahmed Nabih Zaki Rashed, Menoufia University, Egypt
Prof. Shishir K. Shandilya, Rukmani Devi Institute of Science & Technology, India
Dr. J. Komala Lakshmi, SNR Sons College, Computer Science, India
Dr. Muhammad Sohail, KUST, Pakistan
Dr. Manjaiah D.H, Mangalore University, India
Dr. S Santhosh Baboo, D.G.Vaishnav College, Chennai, India
Prof. Dr. Mokhtar Beldjehem, Sainte-Anne University, Halifax, NS, Canada
Dr. Deepak Laxmi Narasimha, University of Malaya, Malaysia
Prof. Dr. Arunkumar Thangavelu, Vellore Institute Of Technology, India
Dr. M. Azath, Anna University, India
Dr. Md. Rabiul Islam, Rajshahi University of Engineering & Technology (RUET), Bangladesh
Dr. Aas Alaa Zaidan Ansaef, Multimedia University, Malaysia
Dr. Suresh Jain, Devi Ahilya University, Indore (MP) India,
Dr. Mohammed M. Kadhum, Universiti Utara Malaysia
Dr. Hanumanthappa. J. University of Mysore, India
Dr. Syed Ishtiaque Ahmed, Bangladesh University of Engineering and Technology (BUET)
Dr. Akinola Solomon Olalekan, University of Ibadan, Ibadan, Nigeria
Dr. Santosh K. Pandey, The Institute of Chartered Accountants of India
Dr. P. Vasant, Power Control Optimization, Malaysia
Dr. Petr Ivankov, Automatika - S, Russian Federation
Dr. Utkarsh Seetha, Data Infosys Limited, India
Mrs. Priti Maheshwary, Maulana Azad National Institute of Technology, Bhopal
Dr. (Mrs) Padmavathi Ganapathi, Avinashilingam University for Women, Coimbatore
Assist. Prof. A. Neela madheswari, Anna university, India
Prof. Ganesan Ramachandra Rao, PSG College of Arts and Science, India
Mr. Kamanashis Biswas, Daffodil International University, Bangladesh
Dr. Atul Gonsai, Saurashtra University, Gujarat, India
Mr. Angkoon Phinyomark, Prince of Songkla University, Thailand
Mrs. G. Nalini Priya, Anna University, Chennai
Dr. P. Subashini, Avinashilingam University for Women, India
Assoc. Prof. Vijay Kumar Chakka, Dhirubhai Ambani IICT, Gandhinagar ,Gujarat
Mr. Jitendra Agrawal, : Rajiv Gandhi Proudhyogiki Vishwavidyalaya, Bhopal
Mr. Vishal Goyal, Department of Computer Science, Punjabi University, India
Dr. R. Baskaran, Department of Computer Science and Engineering, Anna University, Chennai

Assist. Prof. Kanwalvir Singh Dhindsa, B.B.S.B.Engg.College, Fatehgarh Sahib (Punjab), India
Dr. Jamal Ahmad Dargham, School of Engineering and Information Technology, Universiti Malaysia Sabah
Mr. Nitin Bhatia, DAV College, India
Dr. Dhavachelvan Ponnurangam, Pondicherry Central University, India
Dr. Mohd Faizal Abdollah, University of Technical Malaysia, Malaysia
Assist. Prof. Sonal Chawla, Panjab University, India
Dr. Abdul Wahid, AKG Engg. College, Ghaziabad, India
Mr. Arash Habibi Lashkari, University of Malaya (UM), Malaysia
Mr. Md. Rajibul Islam, Ibnu Sina Institute, University Technology Malaysia
Professor Dr. Sabu M. Thampi, .B.S Institute of Technology for Women, Kerala University, India
Mr. Noor Muhammed Nayeem, Université Lumière Lyon 2, 69007 Lyon, France
Dr. Himanshu Aggarwal, Department of Computer Engineering, Punjabi University, India
Prof R. Naidoo, Dept of Mathematics/Center for Advanced Computer Modelling, Durban University of Technology,
Durban, South Africa
Prof. Mydhili K Nair, Visweswaraiah Technological University, Bangalore, India
M. Prabu, Adhiyamaan College of Engineering/Anna University, India
Mr. Swakkhar Shatabda, United International University, Bangladesh
Dr. Abdur Rashid Khan, ICIT, Gomal University, Dera Ismail Khan, Pakistan
Mr. H. Abdul Shabeer, I-Nautix Technologies, Chennai, India
Dr. M. Aramudhan, Perunthalaivar Kamarajar Institute of Engineering and Technology, India
Dr. M. P. Thapliyal, Department of Computer Science, HNB Garhwal University (Central University), India
Dr. Shahaboddin Shamshirband, Islamic Azad University, Iran
Mr. Zeashan Hameed Khan, Université de Grenoble, France
Prof. Anil K Ahlawat, Ajay Kumar Garg Engineering College, Ghaziabad, UP Technical University, Lucknow
Mr. Longe Olumide Babatope, University Of Ibadan, Nigeria
Associate Prof. Raman Maini, University College of Engineering, Punjabi University, India
Dr. Maslin Masrom, University Technology Malaysia, Malaysia
Sudipta Chattopadhyay, Jadavpur University, Kolkata, India
Dr. Dang Tuan NGUYEN, University of Information Technology, Vietnam National University - Ho Chi Minh City
Dr. Mary Lourde R., BITS-PILANI Dubai , UAE
Dr. Abdul Aziz, University of Central Punjab, Pakistan
Mr. Karan Singh, Gautam Budtha University, India
Mr. Avinash Pokhriyal, Uttar Pradesh Technical University, Lucknow, India
Associate Prof Dr Zuraini Ismail, University Technology Malaysia, Malaysia
Assistant Prof. Yasser M. Alginahi, Taibah University, Madinah Munawwarah, KSA
Mr. Dakshina Ranjan Kisku, West Bengal University of Technology, India
Mr. Raman Kumar, Dr B R Ambedkar National Institute of Technology, Jalandhar, Punjab, India
Associate Prof. Samir B. Patel, Institute of Technology, Nirma University, India
Dr. M.Munir Ahamed Rabbani, B. S. Abdur Rahman University, India
Asst. Prof. Koushik Majumder, West Bengal University of Technology, India
Dr. Alex Pappachen James, Queensland Micro-nanotechnology center, Griffith University, Australia
Assistant Prof. S. Hariharan, B.S. Abdur Rahman University, India
Asst Prof. Jasmine. K. S, R.V.College of Engineering, India
Mr Naushad Ali Mamode Khan, Ministry of Education and Human Resources, Mauritius
Prof. Mahesh Goyani, G H Patel Collge of Engg. & Tech, V.V.N, Anand, Gujarat, India
Dr. Mana Mohammed, University of Tlemcen, Algeria
Prof. Jatinder Singh, Universal Institution of Engg. & Tech. CHD, India

Mrs. M. Anandhavalli Gauthaman, Sikkim Manipal Institute of Technology, Majitar, East Sikkim
Dr. Bin Guo, Institute Telecom SudParis, France
Mrs. Maleika Mehr Nigar Mohamed Heenaye-Mamode Khan, University of Mauritius
Prof. Pijush Biswas, RCC Institute of Information Technology, India
Mr. V. Bala Dhandayuthapani, Mekelle University, Ethiopia
Dr. Irfan Syamsuddin, State Polytechnic of Ujung Pandang, Indonesia
Mr. Kavi Kumar Khedo, University of Mauritius, Mauritius
Mr. Ravi Chandiran, Zagro Singapore Pte Ltd. Singapore
Mr. Milindkumar V. Sarode, Jawaharlal Darda Institute of Engineering and Technology, India
Dr. Shamimul Qamar, KSJ Institute of Engineering & Technology, India
Dr. C. Arun, Anna University, India
Assist. Prof. M.N.Birje, Basaveshwar Engineering College, India
Prof. Hamid Reza Naji, Department of Computer Enigneering, Shahid Beheshti University, Tehran, Iran
Assist. Prof. Debasis Giri, Department of Computer Science and Engineering, Haldia Institute of Technology
Subhabrata Barman, Haldia Institute of Technology, West Bengal
Mr. M. I. Lali, COMSATS Institute of Information Technology, Islamabad, Pakistan
Dr. Feroz Khan, Central Institute of Medicinal and Aromatic Plants, Lucknow, India
Mr. R. Nagendran, Institute of Technology, Coimbatore, Tamilnadu, India
Mr. Amnach Khawne, King Mongkut's Institute of Technology Ladkrabang, Ladkrabang, Bangkok, Thailand
Dr. P. Chakrabarti, Sir Padampat Singhanian University, Udaipur, India
Mr. Nafiz Imtiaz Bin Hamid, Islamic University of Technology (IUT), Bangladesh.
Shahab-A. Shamshirband, Islamic Azad University, Chalous, Iran
Prof. B. Priestly Shan, Anna Univeristy, Tamilnadu, India
Venkatramreddy Velma, Dept. of Bioinformatics, University of Mississippi Medical Center, Jackson MS USA
Akshi Kumar, Dept. of Computer Engineering, Delhi Technological University, India
Dr. Umesh Kumar Singh, Vikram University, Ujjain, India
Mr. Serguei A. Mokhov, Concordia University, Canada
Mr. Lai Khin Wee, Universiti Teknologi Malaysia, Malaysia
Dr. Awadhesh Kumar Sharma, Madan Mohan Malviya Engineering College, India
Mr. Syed R. Rizvi, Analytical Services & Materials, Inc., USA
Dr. S. Karthik, SNS College of Technology, India
Mr. Syed Qasim Bukhari, CIMET (Universidad de Granada), Spain
Mr. A.D.Potgantwar, Pune University, India
Dr. Himanshu Aggarwal, Punjabi University, India
Mr. Rajesh Ramachandran, Naipunya Institute of Management and Information Technology, India
Dr. K.L. Shunmuganathan, R.M.K Engg College , Kavaraipettai ,Chennai
Dr. Prasant Kumar Pattnaik, KIST, India.
Dr. Ch. Aswani Kumar, VIT University, India
Mr. Ijaz Ali Shoukat, King Saud University, Riyadh KSA
Mr. Arun Kumar, Sir Padam Pat Singhanian University, Udaipur, Rajasthan
Mr. Muhammad Imran Khan, Universiti Teknologi PETRONAS, Malaysia
Dr. Natarajan Meghanathan, Jackson State University, Jackson, MS, USA
Mr. Mohd Zaki Bin Mas'ud, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia
Prof. Dr. R. Geetharamani, Dept. of Computer Science and Eng., Rajalakshmi Engineering College, India
Dr. Smita Rajpal, Institute of Technology and Management, Gurgaon, India
Dr. S. Abdul Khader Jilani, University of Tabuk, Tabuk, Saudi Arabia
Mr. Syed Jamal Haider Zaidi, Bahria University, Pakistan

Dr. N. Devarajan, Government College of Technology, Coimbatore, Tamilnadu, INDIA
Mr. R. Jagadeesh Kannan, RMK Engineering College, India
Mr. Deo Prakash, Shri Mata Vaishno Devi University, India
Mr. Mohammad Abu Naser, Dept. of EEE, IUT, Gazipur, Bangladesh
Assist. Prof. Prasun Ghosal, Bengal Engineering and Science University, India
Mr. Md. Golam Kaosar, School of Engineering and Science, Victoria University, Melbourne City, Australia
Mr. R. Mahammad Shafi, Madanapalle Institute of Technology & Science, India
Dr. F. Sagayaraj Francis, Pondicherry Engineering College, India
Dr. Ajay Goel, HIET, Kaithal, India
Mr. Nayak Sunil Kashibarao, Bahirji Smarak Mahavidyalaya, India
Mr. Suhas J Manangi, Microsoft India
Dr. Kalyankar N. V., Yeshwant Mahavidyalaya, Nanded, India
Dr. K.D. Verma, S.V. College of Post graduate studies & Research, India
Dr. Amjad Rehman, University Technology Malaysia, Malaysia
Mr. Rachit Garg, L K College, Jalandhar, Punjab
Mr. J. William, M.A.M college of Engineering, Trichy, Tamilnadu, India
Prof. Jue-Sam Chou, Nanhua University, College of Science and Technology, Taiwan
Dr. Thorat S.B., Institute of Technology and Management, India
Mr. Ajay Prasad, Sir Padampat Singhania University, Udaipur, India
Dr. Kamaljit I. Lakhtaria, Atmiya Institute of Technology & Science, India
Mr. Syed Rafiul Hussain, Ahsanullah University of Science and Technology, Bangladesh
Mrs Fazeela Tunnisa, Najran University, Kingdom of Saudi Arabia
Mrs Kavita Taneja, Maharishi Markandeshwar University, Haryana, India
Mr. Maniyar Shiraz Ahmed, Najran University, Najran, KSA
Mr. Anand Kumar, AMC Engineering College, Bangalore
Dr. Rakesh Chandra Gangwar, Beant College of Engg. & Tech., Gurdaspur (Punjab) India
Dr. V V Rama Prasad, Sree Vidyanikethan Engineering College, India
Assist. Prof. Neetesh Kumar Gupta, Technocrats Institute of Technology, Bhopal (M.P.), India
Mr. Ashish Seth, Uttar Pradesh Technical University, Lucknow, UP India
Dr. V V S S S Balaram, Sreenidhi Institute of Science and Technology, India
Mr Rahul Bhatia, Lingaya's Institute of Management and Technology, India
Prof. Niranjana Reddy, P, KITS, Warangal, India
Prof. Rakesh. Lingappa, Vijetha Institute of Technology, Bangalore, India
Dr. Mohammed Ali Hussain, Nimra College of Engineering & Technology, Vijayawada, A.P., India
Dr. A. Srinivasan, MNM Jain Engineering College, Rajiv Gandhi Salai, Thorapakkam, Chennai
Mr. Rakesh Kumar, M.M. University, Mullana, Ambala, India
Dr. Lena Khaled, Zarqa Private University, Aman, Jordan
Ms. Supriya Kapoor, Patni/Lingaya's Institute of Management and Tech., India
Dr. Tossapon Boongoen, Aberystwyth University, UK
Dr. Bilal Alatas, Firat University, Turkey
Assist. Prof. Jyoti Praakash Singh, Academy of Technology, India
Dr. Ritu Soni, GNG College, India
Dr. Mahendra Kumar, Sagar Institute of Research & Technology, Bhopal, India.
Dr. Binod Kumar, Lakshmi Narayan College of Tech. (LNCT) Bhopal India
Dr. Muzhir Shaban Al-Ani, Amman Arab University Amman – Jordan
Dr. T.C. Manjunath, ATRIA Institute of Tech, India
Mr. Muhammad Zakarya, COMSATS Institute of Information Technology (CIIT), Pakistan

Assist. Prof. Harmunish Taneja, M. M. University, India
Dr. Chitra Dhawale , SICSR, Model Colony, Pune, India
Mrs Sankari Muthukaruppan, Nehru Institute of Engineering and Technology, Anna University, India
Mr. Aaqif Afzaal Abbasi, National University Of Sciences And Technology, Islamabad
Prof. Ashutosh Kumar Dubey, Trinity Institute of Technology and Research Bhopal, India
Mr. G. Appasami, Dr. Pauls Engineering College, India
Mr. M Yasin, National University of Science and Tech, karachi (NUST), Pakistan
Mr. Yaser Miaji, University Utara Malaysia, Malaysia
Mr. Shah Ahsanul Haque, International Islamic University Chittagong (IIUC), Bangladesh
Prof. (Dr) Syed Abdul Sattar, Royal Institute of Technology & Science, India
Dr. S. Sasikumar, Roever Engineering College
Assist. Prof. Monit Kapoor, Maharishi Markandeshwar University, India
Mr. Nwaocha Vivian O, National Open University of Nigeria
Dr. M. S. Vijaya, GR Govindarajulu School of Applied Computer Technology, India
Assist. Prof. Chakresh Kumar, Manav Rachna International University, India
Mr. Kunal Chadha , R&D Software Engineer, Gemalto, Singapore
Mr. Mueen Uddin, Universiti Teknologi Malaysia, UTM , Malaysia
Dr. Dhuha Basheer abdullah, Mosul university, Iraq
Mr. S. Audithan, Annamalai University, India
Prof. Vijay K Chaudhari, Technocrats Institute of Technology , India
Associate Prof. Mohd Ilyas Khan, Technocrats Institute of Technology , India
Dr. Vu Thanh Nguyen, University of Information Technology, HoChiMinh City, VietNam
Assist. Prof. Anand Sharma, MITS, Lakshmangarh, Sikar, Rajasthan, India
Prof. T V Narayana Rao, HITAM Engineering college, Hyderabad
Mr. Deepak Gour, Sir Padampat Singhanian University, India
Assist. Prof. Amutharaj Joyson, Kalasalingam University, India
Mr. Ali Balador, Islamic Azad University, Iran
Mr. Mohit Jain, Maharaja Surajmal Institute of Technology, India
Mr. Dilip Kumar Sharma, GLA Institute of Technology & Management, India
Dr. Debojyoti Mitra, Sir padampat Singhanian University, India
Dr. Ali Dehghantanha, Asia-Pacific University College of Technology and Innovation, Malaysia
Mr. Zhao Zhang, City University of Hong Kong, China
Prof. S.P. Setty, A.U. College of Engineering, India
Prof. Patel Rakeshkumar Kantilal, Sankalchand Patel College of Engineering, India
Mr. Biswajit Bhowmik, Bengal College of Engineering & Technology, India
Mr. Manoj Gupta, Apex Institute of Engineering & Technology, India
Assist. Prof. Ajay Sharma, Raj Kumar Goel Institute Of Technology, India
Assist. Prof. Ramveer Singh, Raj Kumar Goel Institute of Technology, India
Dr. Hanan Elazhary, Electronics Research Institute, Egypt
Dr. Hosam I. Faiq, USM, Malaysia
Prof. Dipti D. Patil, MAEER's MIT College of Engg. & Tech, Pune, India
Assist. Prof. Devendra Chack, BCT Kumaon engineering College Dwarahat Almora, India
Prof. Manpreet Singh, M. M. Engg. College, M. M. University, India
Assist. Prof. M. Sadiq ali Khan, University of Karachi, Pakistan
Mr. Prasad S. Halgaonkar, MIT - College of Engineering, Pune, India
Dr. Imran Ghani, Universiti Teknologi Malaysia, Malaysia
Prof. Varun Kumar Kakar, Kumaon Engineering College, Dwarahat, India

Assist. Prof. Nisheeth Joshi, Apaji Institute, Banasthali University, Rajasthan, India
Associate Prof. Kunwar S. Vaisla, VCT Kumaon Engineering College, India
Prof Anupam Choudhary, Bhilai School Of Engg., Bhilai (C.G.), India
Mr. Divya Prakash Shrivastava, Al Jabal Al garbi University, Zawya, Libya
Associate Prof. Dr. V. Radha, Avinashilingam Deemed university for women, Coimbatore.
Dr. Kasarapu Ramani, JNT University, Anantapur, India
Dr. Anuraag Awasthi, Jayoti Vidyapeeth Womens University, India
Dr. C G Ravichandran, R V S College of Engineering and Technology, India
Dr. Mohamed A. Deriche, King Fahd University of Petroleum and Minerals, Saudi Arabia
Mr. Abbas Karimi, Universiti Putra Malaysia, Malaysia
Mr. Amit Kumar, Jaypee University of Engg. and Tech., India
Dr. Nikolai Stoianov, Defense Institute, Bulgaria
Assist. Prof. S. Ranichandra, KSR College of Arts and Science, Tiruchencode
Mr. T.K.P. Rajagopal, Diamond Horse International Pvt Ltd, India
Dr. Md. Ekramul Hamid, Rajshahi University, Bangladesh
Mr. Hemanta Kumar Kalita, TATA Consultancy Services (TCS), India
Dr. Messaouda Azzouzi, Ziane Achour University of Djelfa, Algeria
Prof. (Dr.) Juan Jose Martinez Castillo, "Gran Mariscal de Ayacucho" University and Acantelys research Group, Venezuela
Dr. Jatinderkumar R. Saini, Narmada College of Computer Application, India
Dr. Babak Bashari Rad, University Technology of Malaysia, Malaysia
Dr. Nighat Mir, Effat University, Saudi Arabia
Prof. (Dr.) G.M.Nasira, Sasurie College of Engineering, India
Mr. Varun Mittal, Gemalto Pte Ltd, Singapore
Assist. Prof. Mrs P. Banumathi, Kathir College Of Engineering, Coimbatore
Assist. Prof. Quan Yuan, University of Wisconsin-Stevens Point, US
Dr. Pranam Paul, Narula Institute of Technology, Agarpara, West Bengal, India
Assist. Prof. J. Ramkumar, V.L.B Janakiammal college of Arts & Science, India
Mr. P. Sivakumar, Anna university, Chennai, India
Mr. Md. Humayun Kabir Biswas, King Khalid University, Kingdom of Saudi Arabia
Mr. Mayank Singh, J.P. Institute of Engg & Technology, Meerut, India
HJ. Kamaruzaman Jusoff, Universiti Putra Malaysia
Mr. Nikhil Patrick Lobo, CADES, India
Dr. Amit Wason, Rayat-Bahra Institute of Engineering & Boi-Technology, India
Dr. Rajesh Shrivastava, Govt. Benazir Science & Commerce College, Bhopal, India
Assist. Prof. Vishal Bharti, DCE, Gurgaon
Mrs. Sunita Bansal, Birla Institute of Technology & Science, India
Dr. R. Sudhakar, Dr.Mahalingam college of Engineering and Technology, India
Dr. Amit Kumar Garg, Shri Mata Vaishno Devi University, Katra(J&K), India
Assist. Prof. Raj Gaurang Tiwari, AZAD Institute of Engineering and Technology, India
Mr. Hamed Taherdoost, Tehran, Iran
Mr. Amin Daneshmand Malayeri, YRC, IAU, Malayer Branch, Iran
Mr. Shantanu Pal, University of Calcutta, India
Dr. Terry H. Walcott, E-Promag Consultancy Group, United Kingdom
Dr. Ezekiel U OKIKE, University of Ibadan, Nigeria
Mr. P. Mahalingam, Caledonian College of Engineering, Oman
Dr. Mahmoud M. A. Abd Ellatif, Mansoura University, Egypt

Prof. Kunwar S. Vaisla, BCT Kumaon Engineering College, India
Prof. Mahesh H. Panchal, Kalol Institute of Technology & Research Centre, India
Mr. Muhammad Asad, Technical University of Munich, Germany
Mr. AliReza Shams Shafigh, Azad Islamic university, Iran
Prof. S. V. Nagaraj, RMK Engineering College, India
Mr. Ashikali M Hasan, Senior Researcher, CelNet security, India
Dr. Adnan Shahid Khan, University Technology Malaysia, Malaysia
Mr. Prakash Gajanan Burade, Nagpur University/ITM college of engg, Nagpur, India
Dr. Jagdish B.Helonde, Nagpur University/ITM college of engg, Nagpur, India
Professor, Doctor BOUHORMA Mohammed, Univertsity Abdelmalek Essaadi, Morocco
Mr. K. Thirumalaivasan, Pondicherry Engg. College, India
Mr. Umbarkar Anantkumar Janardan, Walchand College of Engineering, India
Mr. Ashish Chaurasia, Gyan Ganga Institute of Technology & Sciences, India
Mr. Sunil Taneja, Kurukshetra University, India
Mr. Fauzi Adi Rafrastara, Dian Nuswantoro University, Indonesia
Dr. Yaduvir Singh, Thapar University, India
Dr. Ioannis V. Koskosas, University of Western Macedonia, Greece
Dr. Vasantha Kalyani David, Avinashilingam University for women, Coimbatore
Dr. Ahmed Mansour Manasrah, Universiti Sains Malaysia, Malaysia
Miss. Nazanin Sadat Kazazi, University Technology Malaysia, Malaysia
Mr. Saeed Rasouli Heikalabad, Islamic Azad University - Tabriz Branch, Iran
Assoc. Prof. Dharendra Mishra, SVKM's NMIMS University, India
Prof. Shapoor Zarei, UAE Inventors Association, UAE
Prof. B.Raja Sarath Kumar, Lenora College of Engineering, India
Dr. Bashir Alam, Jamia millia Islamia, Delhi, India
Prof. Anant J Umbarkar, Walchand College of Engg., India
Assist. Prof. B. Bharathi, Sathyabama University, India
Dr. Fokrul Alom Mazarbhuiya, King Khalid University, Saudi Arabia
Prof. T.S.Jeyali Laseeth, Anna University of Technology, Tirunelveli, India
Dr. M. Balraju, Jawahar Lal Nehru Technological University Hyderabad, India
Dr. Vijayalakshmi M. N., R.V.College of Engineering, Bangalore
Prof. Walid Moudani, Lebanese University, Lebanon
Dr. Saurabh Pal, VBS Purvanchal University, Jaunpur, India
Associate Prof. Suneet Chaudhary, Dehradun Institute of Technology, India
Associate Prof. Dr. Manuj Darbari, BBD University, India
Ms. Prema Selvaraj, K.S.R College of Arts and Science, India
Assist. Prof. Ms.S.Sasikala, KSR College of Arts & Science, India
Mr. Sukhvinder Singh Deora, NC Institute of Computer Sciences, India
Dr. Abhay Bansal, Amity School of Engineering & Technology, India
Ms. Sumita Mishra, Amity School of Engineering and Technology, India
Professor S. Viswanadha Raju, JNT University Hyderabad, India
Mr. Asghar Shahrzad Khashandarag, Islamic Azad University Tabriz Branch, India
Mr. Manoj Sharma, Panipat Institute of Engg. & Technology, India
Mr. Shakeel Ahmed, King Faisal University, Saudi Arabia
Dr. Mohamed Ali Mahjoub, Institute of Engineer of Monastir, Tunisia
Mr. Adri Jovin J.J., SriGuru Institute of Technology, India
Dr. Sukumar Senthilkumar, Universiti Sains Malaysia, Malaysia

Mr. Rakesh Bharati, Dehradun Institute of Technology Dehradun, India
Mr. Shervan Fekri Ershad, Shiraz International University, Iran
Mr. Md. Safiqul Islam, Daffodil International University, Bangladesh
Mr. Mahmudul Hasan, Daffodil International University, Bangladesh
Prof. Mandakini Tayade, UIT, RGTU, Bhopal, India
Ms. Sarla More, UIT, RGTU, Bhopal, India
Mr. Tushar Hrishikesh Jaware, R.C. Patel Institute of Technology, Shirpur, India
Ms. C. Divya, Dr G R Damodaran College of Science, Coimbatore, India
Mr. Fahimuddin Shaik, Annamacharya Institute of Technology & Sciences, India
Dr. M. N. Giri Prasad, JNTUCE, Pulivendula, A.P., India
Assist. Prof. Chintan M Bhatt, Charotar University of Science And Technology, India
Prof. Sahista Machchhar, Marwadi Education Foundation's Group of institutions, India
Assist. Prof. Navnish Goel, S. D. College Of Engineering & Technology, India
Mr. Khaja Kamaluddin, Sirt University, Sirt, Libya
Mr. Mohammad Zaidul Karim, Daffodil International, Bangladesh
Mr. M. Vijayakumar, KSR College of Engineering, Tiruchengode, India
Mr. S. A. Ahsan Rajon, Khulna University, Bangladesh
Dr. Muhammad Mohsin Nazir, LCW University Lahore, Pakistan
Mr. Mohammad Asadul Hoque, University of Alabama, USA
Mr. P.V.Sarathchand, Indur Institute of Engineering and Technology, India
Mr. Durgesh Samadhiya, Chung Hua University, Taiwan
Dr Venu Kuthadi, University of Johannesburg, Johannesburg, RSA
Dr. (Er) Jasvir Singh, Guru Nanak Dev University, Amritsar, Punjab, India
Mr. Jasmin Cosic, Min. of the Interior of Una-sana canton, B&H, Bosnia and Herzegovina
Dr S. Rajalakshmi, Botho College, South Africa
Dr. Mohamed Sarrah, De Montfort University, UK
Mr. Basappa B. Kodada, Canara Engineering College, India
Assist. Prof. K. Ramana, Annamacharya Institute of Technology and Sciences, India
Dr. Ashu Gupta, Apeejay Institute of Management, Jalandhar, India
Assist. Prof. Shaik Rasool, Shadan College of Engineering & Technology, India
Assist. Prof. K. Suresh, Annamacharya Institute of Tech & Sci. Rajampet, AP, India
Dr . G. Singaravel, K.S.R. College of Engineering, India
Dr B. G. Geetha, K.S.R. College of Engineering, India
Assist. Prof. Kavita Choudhary, ITM University, Gurgaon
Dr. Mehrdad Jalali, Azad University, Mashhad, Iran
Megha Goel, Shamli Institute of Engineering and Technology, Shamli, India
Mr. Chi-Hua Chen, Institute of Information Management, National Chiao-Tung University, Taiwan (R.O.C.)
Assoc. Prof. A. Rajendran, RVS College of Engineering and Technology, India
Assist. Prof. S. Jaganathan, RVS College of Engineering and Technology, India
Assoc. Prof. (Dr.) A S N Chakravarthy, JNTUK University College of Engineering Vizianagaram (State University)
Assist. Prof. Deepshikha Patel, Technocrat Institute of Technology, India
Assist. Prof. Maram Balajee, GMRIT, India
Assist. Prof. Monika Bhatnagar, TIT, India
Prof. Gaurang Panchal, Charotar University of Science & Technology, India
Prof. Anand K. Tripathi, Computer Society of India
Prof. Jyoti Chaudhary, High Performance Computing Research Lab, India
Assist. Prof. Supriya Raheja, ITM University, India

Dr. Pankaj Gupta, Microsoft Corporation, U.S.A.
Assist. Prof. Panchamukesh Chandaka, Hyderabad Institute of Tech. & Management, India
Prof. Mohan H.S, SJB Institute Of Technology, India
Mr. Hossein Malekinezhad, Islamic Azad University, Iran
Mr. Zatin Gupta, Universti Malaysia, Malaysia
Assist. Prof. Amit Chauhan, Phonics Group of Institutions, India
Assist. Prof. Ajal A. J., METS School Of Engineering, India
Mrs. Omowunmi Omobola Adeyemo, University of Ibadan, Nigeria
Dr. Bharat Bhushan Agarwal, I.F.T.M. University, India
Md. Nazrul Islam, University of Western Ontario, Canada
Tushar Kanti, L.N.C.T, Bhopal, India
Er. Aumreesh Kumar Saxena, SIRTs College Bhopal, India
Mr. Mohammad Monirul Islam, Daffodil International University, Bangladesh
Dr. Kashif Nisar, University Utara Malaysia, Malaysia
Dr. Wei Zheng, Rutgers Univ/ A10 Networks, USA
Associate Prof. Rituraj Jain, Vyas Institute of Engg & Tech, Jodhpur – Rajasthan
Assist. Prof. Apoorvi Sood, I.T.M. University, India
Dr. Kayhan Zrar Ghafoor, University Technology Malaysia, Malaysia
Mr. Swapnil Sonar, Truba Institute College of Engineering & Technology, Indore, India
Ms. Yogita Gigras, I.T.M. University, India
Associate Prof. Neelima Sadineni, Pydha Engineering College, India Pydha Engineering College
Assist. Prof. K. Deepika Rani, HITAM, Hyderabad
Ms. Shikha Maheshwari, Jaipur Engineering College & Research Centre, India
Prof. Dr V S Giridhar Akula, Avanthi's Scientific Tech. & Research Academy, Hyderabad
Prof. Dr.S.Saravanan, Muthayammal Engineering College, India
Mr. Mehdi Golsorkhatabar Amiri, Islamic Azad University, Iran
Prof. Amit Sadanand Savyanavar, MITCOE, Pune, India
Assist. Prof. P.Oliver Jayaprakash, Anna University, Chennai
Assist. Prof. Ms. Sujata, ITM University, Gurgaon, India
Dr. Asoke Nath, St. Xavier's College, India
Mr. Masoud Rafighi, Islamic Azad University, Iran
Assist. Prof. RamBabu Pemula, NIMRA College of Engineering & Technology, India
Assist. Prof. Ms Rita Chhikara, ITM University, Gurgaon, India
Mr. Sandeep Maan, Government Post Graduate College, India
Prof. Dr. S. Muralidharan, Mepco Schlenk Engineering College, India
Associate Prof. T.V.Sai Krishna, QIS College of Engineering and Technology, India
Mr. R. Balu, Bharathiar University, Coimbatore, India
Assist. Prof. Shekhar. R, Dr.SM College of Engineering, India
Prof. P. Senthilkumar, Vivekanandha Institue of Engineering and Technology for Woman, India
Mr. M. Kamarajan, PSNA College of Engineering & Technology, India
Dr. Angajala Srinivasa Rao, Jawaharlal Nehru Technical University, India
Assist. Prof. C. Venkatesh, A.I.T.S, Rajampet, India
Mr. Afshin Rezakhani Roozbahani, Ayatollah Boroujerdi University, Iran
Mr. Laxmi chand, SCTL, Noida, India
Dr. Dr. Abdul Hannan, Vivekanand College, Aurangabad
Prof. Mahesh Panchal, KITRC, Gujarat
Dr. A. Subramani, K.S.R. College of Engineering, Tiruchengode

Assist. Prof. Prakash M, Rajalakshmi Engineering College, Chennai, India
Assist. Prof. Akhilesh K Sharma, Sir Padampat Singhanian University, India
Ms. Varsha Sahni, Guru Nanak Dev Engineering College, Ludhiana, India
Associate Prof. Trilochan Rout, NM Institute of Engineering and Technology, India
Mr. Srikanta Kumar Mohapatra, NMIET, Orissa, India
Mr. Waqas Haider Bangyal, Iqra University Islamabad, Pakistan
Dr. S. Vijayaragavan, Christ College of Engineering and Technology, Pondicherry, India
Prof. Elbouchari Mohamed, University Mohammed First, Oujda, Morocco
Dr. Muhammad Asif Khan, King Faisal University, Saudi Arabia
Dr. Nagy Ramadan Darwish Omran, Cairo University, Egypt.
Assistant Prof. Anand Nayyar, KCL Institute of Management and Technology, India
Mr. G. Premsankar, Ericsson, India
Assist. Prof. T. Hemalatha, VELS University, India
Prof. Tejaswini Apte, University of Pune, India
Dr. Edmund Ng Giap Weng, Universiti Malaysia Sarawak, Malaysia
Mr. Mahdi Nouri, Iran University of Science and Technology, Iran
Associate Prof. S. Asif Hussain, Annamacharya Institute of technology & Sciences, India
Mrs. Kavita Pabreja, Maharaja Surajmal Institute (an affiliate of GGSIP University), India
Mr. Vorugunti Chandra Sekhar, DA-IICT, India
Mr. Muhammad Najmi Ahmad Zabidi, Universiti Teknologi Malaysia, Malaysia
Dr. Aderemi A. Atayero, Covenant University, Nigeria
Assist. Prof. Osama Sohaib, Balochistan University of Information Technology, Pakistan
Assist. Prof. K. Suresh, Annamacharya Institute of Technology and Sciences, India
Mr. Hassen Mohammed Abdullaah Alsafi, International Islamic University Malaysia (IIUM) Malaysia
Mr. Robail Yasrab, Virtual University of Pakistan, Pakistan
Mr. R. Balu, Bharathiar University, Coimbatore, India
Prof. Anand Nayyar, KCL Institute of Management and Technology, Jalandhar
Assoc. Prof. Vivek S Deshpande, MIT College of Engineering, India
Prof. K. Saravanan, Anna university Coimbatore, India
Dr. Ravendra Singh, MJP Rohilkhand University, Bareilly, India
Mr. V. Mathivanan, IBRA College of Technology, Sultanate of OMAN
Assoc. Prof. S. Asif Hussain, AITS, India
Assist. Prof. C. Venkatesh, AITS, India
Mr. Sami Ulhaq, SZABIST Islamabad, Pakistan
Dr. B. Justus Rabi, Institute of Science & Technology, India
Mr. Anuj Kumar Yadav, Dehradun Institute of technology, India
Mr. Alejandro Mosquera, University of Alicante, Spain
Assist. Prof. Arjun Singh, Sir Padampat Singhanian University (SPSU), Udaipur, India
Dr. Smriti Agrawal, JB Institute of Engineering and Technology, Hyderabad
Assist. Prof. Swathi Sambangi, Visakha Institute of Engineering and Technology, India
Ms. Prabhjot Kaur, Guru Gobind Singh Indraprastha University, India
Mrs. Samaher AL-Hothali, Yanbu University College, Saudi Arabia
Prof. Rajneeshkaur Bedi, MIT College of Engineering, Pune, India
Mr. Hassen Mohammed Abdullaah Alsafi, International Islamic University Malaysia (IIUM)
Dr. Wei Zhang, Amazon.com, Seattle, WA, USA
Mr. B. Santhosh Kumar, C S I College of Engineering, Tamil Nadu
Dr. K. Reji Kumar, N S S College, Pandalam, India

Assoc. Prof. K. Seshadri Sastry, EILM University, India
Mr. Kai Pan, UNC Charlotte, USA
Mr. Ruikar Sachin, SGGSIET, India
Prof. (Dr.) Vinodani Katiyar, Sri Ramswaroop Memorial University, India
Assoc. Prof., M. Giri, Sreenivasa Institute of Technology and Management Studies, India
Assoc. Prof. Labib Francis Gergis, Misr Academy for Engineering and Technology (MET), Egypt
Assist. Prof. Amanpreet Kaur, ITM University, India
Assist. Prof. Anand Singh Rajawat, Shri Vaishnav Institute of Technology & Science, Indore
Mrs. Hadeel Saleh Haj Aliwi, Universiti Sains Malaysia (USM), Malaysia
Dr. Abhay Bansal, Amity University, India
Dr. Mohammad A. Mezher, Fahad Bin Sultan University, KSA
Assist. Prof. Nidhi Arora, M.C.A. Institute, India
Prof. Dr. P. Suresh, Karpagam College of Engineering, Coimbatore, India
Dr. Kannan Balasubramanian, Mepco Schlenk Engineering College, India
Dr. S. Sankara Gomathi, Panimalar Engineering college, India
Prof. Anil kumar Suthar, Gujarat Technological University, L.C. Institute of Technology, India
Assist. Prof. R. Hubert Rajan, NOORUL ISLAM UNIVERSITY, India
Assist. Prof. Dr. Jyoti Mahajan, College of Engineering & Technology
Assist. Prof. Homam Reda El-Taj, College of Network Engineering, Saudi Arabia & Malaysia
Mr. Bijan Paul, Shahjalal University of Science & Technology, Bangladesh
Assoc. Prof. Dr. Ch V Phani Krishna, KL University, India
Dr. Vishal Bhatnagar, Ambedkar Institute of Advanced Communication Technologies & Research, India
Dr. Lamri LAOUAMER, Al Qassim University, Dept. Info. Systems & European University of Brittany, Dept. Computer Science, UBO, Brest, France
Prof. Ashish Babanrao Sasankar, G.H.Raisoni Institute Of Information Technology, India
Prof. Pawan Kumar Goel, Shamli Institute of Engineering and Technology, India
Mr. Ram Kumar Singh, S.V Subharti University, India
Assistant Prof. Sunish Kumar O S, Amaljiyothi College of Engineering, India
Dr Sanjay Bhargava, Banasthali University, India
Mr. Pankaj S. Kulkarni, AVEW's Shatabdi Institute of Technology, India
Mr. Roohollah Etemadi, Islamic Azad University, Iran
Mr. Oloruntoyin Sefiu Taiwo, Emmanuel Alayande College Of Education, Nigeria
Mr. Sumit Goyal, National Dairy Research Institute, India
Mr Jaswinder Singh Dilawari, Geeta Engineering College, India
Prof. Raghuraj Singh, Harcourt Butler Technological Institute, Kanpur
Dr. S.K. Mahendran, Anna University, Chennai, India
Dr. Amit Wason, Hindustan Institute of Technology & Management, Punjab
Dr. Ashu Gupta, Apeejay Institute of Management, India
Assist. Prof. D. Asir Antony Gnana Singh, M.I.E.T Engineering College, India
Mrs Mina Farmanbar, Eastern Mediterranean University, Famagusta, North Cyprus
Mr. Maram Balajee, GMR Institute of Technology, India
Mr. Moiz S. Ansari, Isra University, Hyderabad, Pakistan
Mr. Adebayo, Olawale Surajudeen, Federal University of Technology Minna, Nigeria
Mr. Jasvir Singh, University College Of Engg., India
Mr. Vivek Tiwari, MANIT, Bhopal, India
Assoc. Prof. R. Navaneethakrishnan, Bharathiyar College of Engineering and Technology, India
Mr. Somdip Dey, St. Xavier's College, Kolkata, India

Mr. Souleymane Balla-Arabé, Xi'an University of Electronic Science and Technology, China
Mr. Mahabub Alam, Rajshahi University of Engineering and Technology, Bangladesh
Mr. Sathyapraksh P., S.K.P Engineering College, India
Dr. N. Karthikeyan, SNS College of Engineering, Anna University, India
Dr. Binod Kumar, JSPM's, Jayawant Technical Campus, Pune, India
Assoc. Prof. Dinesh Goyal, Suresh Gyan Vihar University, India
Mr. Md. Abdul Ahad, K L University, India
Mr. Vikas Bajpai, The LNM IIT, India
Dr. Manish Kumar Anand, Salesforce (R & D Analytics), San Francisco, USA
Assist. Prof. Dheeraj Murari, Kumaon Engineering College, India
Assoc. Prof. Dr. A. Muthukumaravel, VELS University, Chennai
Mr. A. Siles Balasingh, St. Joseph University in Tanzania, Tanzania
Mr. Ravindra Daga Badgujar, R C Patel Institute of Technology, India
Dr. Preeti Khanna, SVKM's NMIMS, School of Business Management, India
Mr. Kumar Dayanand, Cambridge Institute of Technology, India
Dr. Syed Asif Ali, SMI University Karachi, Pakistan
Prof. Pallvi Pandit, Himachal Pradesh University, India
Mr. Ricardo Verschueren, University of Gloucestershire, UK
Assist. Prof. Mamta Juneja, University Institute of Engineering and Technology, Panjab University, India
Assoc. Prof. P. Surendra Varma, NRI Institute of Technology, JNTU Kakinada, India
Assist. Prof. Gaurav Shrivastava, RGPV / SVITS Indore, India
Dr. S. Sumathi, Anna University, India
Assist. Prof. Ankita M. Kapadia, Charotar University of Science and Technology, India
Mr. Deepak Kumar, Indian Institute of Technology (BHU), India
Dr. Dr. Rajan Gupta, GGSIP University, New Delhi, India
Assist. Prof. M. Anand Kumar, Karpagam University, Coimbatore, India
Mr. Mr Arshad Mansoor, Pakistan Aeronautical Complex
Mr. Kapil Kumar Gupta, Ansal Institute of Technology and Management, India
Dr. Neeraj Tomer, SINE International Institute of Technology, Jaipur, India
Assist. Prof. Trunal J. Patel, C.G.Patel Institute of Technology, Uka Tarsadia University, Bardoli, Surat
Mr. Sivakumar, Codework solutions, India
Mr. Mohammad Sadegh Mirzaei, PGNR Company, Iran
Dr. Gerard G. Dumancas, Oklahoma Medical Research Foundation, USA
Mr. Varadala Sridhar, Varadhaman College Engineering College, Affiliated To JNTU, Hyderabad
Assist. Prof. Manoj Dhawan, SVITS, Indore
Assoc. Prof. Chitreshh Banerjee, Suresh Gyan Vihar University, Jaipur, India
Dr. S. Santhi, SCSVMV University, India
Mr. Davood Mohammadi Souran, Ministry of Energy of Iran, Iran
Mr. Shamim Ahmed, Bangladesh University of Business and Technology, Bangladesh
Mr. Sandeep Reddivari, Mississippi State University, USA
Assoc. Prof. Ousmane Thiare, Gaston Berger University, Senegal
Dr. Hazra Imran, Athabasca University, Canada
Dr. Setu Kumar Chaturvedi, Technocrats Institute of Technology, Bhopal, India
Mr. Mohd Dilshad Ansari, Jaypee University of Information Technology, India
Ms. Jaspreet Kaur, Distance Education LPU, India
Dr. D. Nagarajan, Salalah College of Technology, Sultanate of Oman
Dr. K.V.N.R.Sai Krishna, S.V.R.M. College, India

Mr. Himanshu Pareek, Center for Development of Advanced Computing (CDAC), India
Mr. Khaldi Amine, Badji Mokhtar University, Algeria
Mr. Mohammad Sadegh Mirzaei, Scientific Applied University, Iran
Assist. Prof. Khyati Chaudhary, Ram-eesh Institute of Engg. & Technology, India
Mr. Sanjay Agal, Pacific College of Engineering Udaipur, India
Mr. Abdul Mateen Ansari, King Khalid University, Saudi Arabia
Dr. H.S. Behera, Veer Surendra Sai University of Technology (VSSUT), India
Dr. Shrikant Tiwari, Shri Shankaracharya Group of Institutions (SSGI), India
Prof. Ganesh B. Regulwar, Shri Shankarprasad Agnihotri College of Engg, India
Prof. Pinnamaneni Bhanu Prasad, Matrix vision GmbH, Germany
Dr. Shrikant Tiwari, Shri Shankaracharya Technical Campus (SSTC), India
Dr. Siddesh G.K., : Dayananada Sagar College of Engineering, Bangalore, India
Dr. Nadir Bouchama, CERIST Research Center, Algeria
Dr. R. Sathishkumar, Sri Venkateswara College of Engineering, India
Assistant Prof (Dr.) Mohamed Moussaoui, Abdelmalek Essaadi University, Morocco
Dr. S. Malathi, Panimalar Engineering College, Chennai, India
Dr. V. Subedha, Panimalar Institute of Technology, Chennai, India
Dr. Prashant Panse, Swami Vivekanand College of Engineering, Indore, India
Dr. Hamza Aldabbas, Al-Balqa'a Applied University, Jordan
Dr. G. Rasitha Banu, Vel's University, Chennai
Dr. V. D. Ambeth Kumar, Panimalar Engineering College, Chennai
Prof. Anuranjan Misra, Bhagwant Institute of Technology, Ghaziabad, India
Ms. U. Sinthuja, PSG college of arts & science, India
Dr. Ehsan Saradar Torshizi, Urmia University, Iran
Dr. Shamneesh Sharma, APG Shimla University, Shimla (H.P.), India
Assistant Prof. A. S. Syed Navaz, Muthayammal College of Arts & Science, India
Assistant Prof. Ranjit Panigrahi, Sikkim Manipal Institute of Technology, Majitar, Sikkim
Dr. Khaled Eskaf, Arab Academy for Science ,Technology & Maritime Transportation, Egypt
Dr. Nishant Gupta, University of Jammu, India
Assistant Prof. Nagarajan Sankaran, Annamalai University, Chidambaram, Tamilnadu, India
Assistant Prof. Tribikram Pradhan, Manipal Institute of Technology, India
Dr. Nasser Lotfi, Eastern Mediterranean University, Northern Cyprus
Dr. R. Manavalan, K S Rangasamy college of Arts and Science, Tamilnadu, India
Assistant Prof. P. Krishna Sankar, K S Rangasamy college of Arts and Science, Tamilnadu, India
Dr. Rahul Malik, Cisco Systems, USA
Dr. S. C. Lingareddy, ALPHA College of Engineering, India
Assistant Prof. Mohammed Shuaib, Interat University, Lucknow, India
Dr. Sachin Yele, Sanghvi Institute of Management & Science, India
Dr. T. Thambidurai, Sun Univercell, Singapore
Prof. Anandkumar Telang, BKIT, India
Assistant Prof. R. Poorvadevi, SCSVMV University, India
Dr Uttam Mande, Gitam University, India
Dr. Poornima Girish Naik, Shahu Institute of Business Education and Research (SIBER), India
Prof. Md. Abu Kausar, Jaipur National University, Jaipur, India
Dr. Mohammed Zuber, AISECT University, India
Prof. Kalum Priyanath Udagepola, King Abdulaziz University, Saudi Arabia
Dr. K. R. Ananth, Velalar College of Engineering and Technology, India

Assistant Prof. Sanjay Sharma, Roorkee Engineering & Management Institute Shamli (U.P), India
Assistant Prof. Panem Charan Arur, Priyadarshini Institute of Technology, India
Dr. Ashwak Mahmood muhsen alabaichi, Karbala University / College of Science, Iraq
Dr. Urmila Shrawankar, G H Raison College of Engineering, Nagpur (MS), India
Dr. Krishan Kumar Paliwal, Panipat Institute of Engineering & Technology, India
Dr. Mukesh Negi, Tech Mahindra, India
Dr. Anuj Kumar Singh, Amity University Gurgaon, India
Dr. Babar Shah, Gyeongsang National University, South Korea
Assistant Prof. Jayprakash Upadhyay, SRI-TECH Jabalpur, India
Assistant Prof. Varadala Sridhar, Vidya Jyothi Institute of Technology, India
Assistant Prof. Parameshachari B D, KSIT, Bangalore, India
Assistant Prof. Ankit Garg, Amity University, Haryana, India
Assistant Prof. Rajashe Karappa, SDMCET, Karnataka, India
Assistant Prof. Varun Jasuja, GNIT, India
Assistant Prof. Sonal Honale, Abha Gaikwad Patil College of Engineering Nagpur, India
Dr. Pooja Choudhary, CT Group of Institutions, NIT Jalandhar, India
Dr. Faouzi Hidoussi, UHL Batna, Algeria
Dr. Naseer Ali Hussein, Wasit University, Iraq
Assistant Prof. Vinod Kumar Shukla, Amity University, Dubai
Dr. Ahmed Farouk Metwaly, K L University
Mr. Mohammed Noaman Murad, Cihan University, Iraq
Dr. Suxing Liu, Arkansas State University, USA
Dr. M. Gomathi, Velalar College of Engineering and Technology, India
Assistant Prof. Sumardiono, College PGRI Blitar, Indonesia
Dr. Latika Kharb, Jagan Institute of Management Studies (JIMS), Delhi, India
Associate Prof. S. Raja, Pauls College of Engineering and Technology, Tamilnadu, India
Assistant Prof. Seyed Reza Pakize, Shahid Sani High School, Iran
Dr. Thiyagu Nagaraj, University-INOUE, India
Assistant Prof. Noreen Sarai, Harare Institute of Technology, Zimbabwe
Assistant Prof. Gajanand Sharma, Suresh Gyan Vihar University Jaipur, Rajasthan, India
Assistant Prof. Mapari Vikas Prakash, Siddhant COE, Sudumbare, Pune, India
Dr. Devesh Katiyar, Shri Ramswaroop Memorial University, India
Dr. Shenshen Liang, University of California, Santa Cruz, US
Assistant Prof. Mohammad Abu Omar, Limkokwing University of Creative Technology- Malaysia
Mr. Snehasis Banerjee, Tata Consultancy Services, India
Assistant Prof. Kibona Lusekelo, Ruaha Catholic University (RUCU), Tanzania
Assistant Prof. Adib Kabir Chowdhury, University College Technology Sarawak, Malaysia
Dr. Ying Yang, Computer Science Department, Yale University, USA
Dr. Vinay Shukla, Institute Of Technology & Management, India
Dr. Liviu Octavian Maftciu-Scai, West University of Timisoara, Romania
Assistant Prof. Rana Khudhair Abbas Ahmed, Al-Rafidain University College, Iraq
Assistant Prof. Nitin A. Naik, S.R.T.M. University, India
Dr. Timothy Powers, University of Hertfordshire, UK
Dr. S. Prasath, Bharathiar University, Erode, India
Dr. Ritu Shrivastava, SIRTIS Bhopal, India
Prof. Rohit Shrivastava, Mittal Institute of Technology, Bhopal, India
Dr. Gianina Mihai, Dunarea de Jos" University of Galati, Romania

Assistant Prof. Ms. T. Kalai Selvi, Erode Sengunthar Engineering College, India
Assistant Prof. Ms. C. Kavitha, Erode Sengunthar Engineering College, India
Assistant Prof. K. Sinivasamoorthi, Erode Sengunthar Engineering College, India
Assistant Prof. Mallikarjun C Sarsamba Bheemna Khandre Institute Technology, Bhalki, India
Assistant Prof. Vishwanath Chikaraddi, Veermata Jijabai technological Institute (Central Technological Institute), India
Assistant Prof. Dr. Ikvinderpal Singh, Trai Shatabdi GGS Khalsa College, India
Assistant Prof. Mohammed Noaman Murad, Cihan University, Iraq
Professor Yousef Farhaoui, Moulay Ismail University, Errachidia, Morocco
Dr. Parul Verma, Amity University, India
Professor Yousef Farhaoui, Moulay Ismail University, Errachidia, Morocco
Assistant Prof. Madhavi Dhingra, Amity University, Madhya Pradesh, India
Assistant Prof.. G. Selvavinayagam, SNS College of Technology, Coimbatore, India
Assistant Prof. Madhavi Dhingra, Amity University, MP, India
Professor Kartheesan Log, Anna University, Chennai
Professor Vasudeva Acharya, Shri Madhwa vadiraja Institute of Technology, India
Dr. Asif Iqbal Hajamydeen, Management & Science University, Malaysia
Assistant Prof., Mahendra Singh Meena, Amity University Haryana
Assistant Professor Manjeet Kaur, Amity University Haryana
Dr. Mohamed Abd El-Basset Matwalli, Zagazig University, Egypt
Dr. Ramani Kannan, Universiti Teknologi PETRONAS, Malaysia
Assistant Prof. S. Jagadeesan Subramaniam, Anna University, India
Assistant Prof. Dharmendra Choudhary, Tripura University, India
Assistant Prof. Deepika Vodnala, SR Engineering College, India
Dr. Kai Cong, Intel Corporation & Computer Science Department, Portland State University, USA
Dr. Kailas R Patil, Vishwakarma Institute of Information Technology (VIIT), India
Dr. Omar A. Alzubi, Faculty of IT / Al-Balqa Applied University, Jordan
Assistant Prof. Kareemullah Shaik, Nimra Institute of Science and Technology, India
Assistant Prof. Chirag Modi, NIT Goa
Dr. R. Ramkumar, Nandha Arts And Science College, India
Dr. Priyadarshini Vydhialingam, Harathiar University, India
Dr. P. S. Jagadeesh Kumar, DBIT, Bangalore, Karnataka
Dr. Vikas Thada, AMITY University, Pachgaon
Dr. T. A. Ashok Kumar, Institute of Management, Christ University, Bangalore
Dr. Shaheera Rashwan, Informatics Research Institute
Dr. S. Preetha Gunasekar, Bharathiyar University, India
Asst Professor Sameer Dev Sharma, Uttaranchal University, Dehradun
Dr. Zhihan Iv, Chinese Academy of Science, China
Dr. Ikvinderpal Singh, Trai Shatabdi GGS Khalsa College, Amritsar
Dr. Umar Ruhi, University of Ottawa, Canada
Dr. Jasmin Cosic, University of Bihac, Bosnia and Herzegovina
Dr. Homam Reda El-Taj, University of Tabuk, Kingdom of Saudi Arabia
Dr. Mostafa Ghobaei Arani, Islamic Azad University, Iran
Dr. Ayyasamy Ayyanar, Annamalai University, India
Dr. Selvakumar Manickam, Universiti Sains Malaysia, Malaysia
Dr. Murali Krishna Namana, GITAM University, India
Dr. Smriti Agrawal, Chaitanya Bharathi Institute of Technology, Hyderabad, India
Professor Vimalathithan Rathinasabapathy, Karpagam College Of Engineering, India

Dr. Sushil Chandra Dimri, Graphic Era University, India
Dr. Dinh-Sinh Mai, Le Quy Don Technical University, Vietnam
Dr. S. Rama Sree, Aditya Engg. College, India
Dr. Ehab T. Alnfwawy, Sadat Academy, Egypt
Dr. Patrick D. Cerna, Haramaya University, Ethiopia
Dr. Vishal Jain, Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), India
Associate Prof. Dr. Jiliang Zhang, North Eastern University, China
Dr. Sharefa Murad, Middle East University, Jordan
Dr. Ajeet Singh Poonia, Govt. College of Engineering & technology, Rajasthan, India
Dr. Vahid Esmaealzadeh, University of Science and Technology, Iran
Dr. Jacek M. Czerniak, Casimir the Great University in Bydgoszcz, Institute of Technology, Poland
Associate Prof. Anisur Rehman Nasir, Jamia Millia Islamia University
Assistant Prof. Imran Ahmad, COMSATS Institute of Information Technology, Pakistan
Professor Ghulam Qasim, Preston University, Islamabad, Pakistan
Dr. Parameshachari B D, GSSS Institute of Engineering and Technology for Women
Dr. Wencan Luo, University of Pittsburgh, US
Dr. Musa PEKER, Faculty of Technology, Mugla Sitki Kocman University, Turkey
Dr. Gunasekaran Shanmugam, Anna University, India
Dr. Binh P. Nguyen, National University of Singapore, Singapore
Dr. Rajkumar Jain, Indian Institute of Technology Indore, India
Dr. Imtiaz Ali Halepoto, QUEST Nawabshah, Pakistan
Dr. Shaligram Prajapat, Devi Ahilya University Indore India
Dr. Sunita Singhal, Birla Institute of Technology and Science, Pilani, India
Dr. Ijaz Ali Shoukat, King Saud University, Saudi Arabia
Dr. Anuj Gupta, IKG Punjab Technical University, India
Dr. Sonali Saini, IES-IPS Academy, India
Dr. Krishan Kumar, Moti Lal Nehru National Institute of Technology, Allahabad, India
Dr. Z. Faizal Khan, College of Engineering, Shaqra University, Kingdom of Saudi Arabia
Prof. M. Padmavathamma, S.V. University Tirupati, India
Prof. A. Velayudham, Cape Institute of Technology, India
Prof. Seifeidne Kadry, American University of the Middle East
Dr. J. Durga Prasad Rao, Pt. Ravishankar Shukla University, Raipur
Assistant Prof. Najam Hasan, Dhofar University
Dr. G. Suseendran, Vels University, Pallavaram, Chennai
Prof. Ankit Faldu, Gujarat Technological University- Atmiya Institute of Technology and Science
Dr. Ali Habiboghli, Islamic Azad University
Dr. Deepak Dembla, JECRC University, Jaipur, India
Dr. Pankaj Rajan, Walmart Labs, USA
Assistant Prof. Radoslava Kraveva, South-West University "Neofit Rilski", Bulgaria
Assistant Prof. Medhavi Shriwas, Shri vaishnav institute of Technology, India
Associate Prof. Sedat Akleylek, Ondokuz Mayıs University, Turkey
Dr. U.V. Arivazhagu, Kingston Engineering College Affiliated To Anna University, India
Dr. Touseef Ali, University of Engineering and Technology, Taxila, Pakistan
Assistant Prof. Naren Jeeva, SASTRA University, India
Dr. Riccardo Colella, University of Salento, Italy
Dr. Enache Maria Cristina, University of Galati, Romania
Dr. Senthil P, Kuringi College of Arts & Science, India

Dr. Hasan Ashrafi-rizi, Isfahan University of Medical Sciences, Isfahan, Iran
Dr. Mazhar Malik, Institute of Southern Punjab, Pakistan
Dr. Yajie Miao, Carnegie Mellon University, USA
Dr. Kamran Shaukat, University of the Punjab, Pakistan
Dr. Sasikaladevi N., SASTRA University, India
Dr. Ali Asghar Rahmani Hosseinabadi, Islamic Azad University Ayatollah Amoli Branch, Amol, Iran
Dr. Velin Kralev, South-West University "Neofit Rilski", Blagoevgrad, Bulgaria
Dr. Marius Iulian Mihailescu, LUMINA - The University of South-East Europe
Dr. Sriramula Nagaprasad, S.R.R.Govt.Arts & Science College, Karimnagar, India
Prof (Dr.) Namrata Dhanda, Dr. APJ Abdul Kalam Technical University, Lucknow, India
Dr. Javed Ahmed Mahar, Shah Abdul Latif University, Khairpur Mir's, Pakistan
Dr. B. Narendra Kumar Rao, Sree Vidyanikethan Engineering College, India
Dr. Shahzad Anwar, University of Engineering & Technology Peshawar, Pakistan
Dr. Basit Shahzad, King Saud University, Riyadh - Saudi Arabia
Dr. Nilamadhab Mishra, Chang Gung University
Dr. Sachin Kumar, Indian Institute of Technology Roorkee
Dr. Santosh Nanda, Biju-Pattnaik University of Technology
Dr. Sherzod Turaev, International Islamic University Malaysia
Dr. Yilun Shang, Tongji University, Department of Mathematics, Shanghai, China
Dr. Nuzhat Shaikh, Modern Education society's College of Engineering, Pune, India
Dr. Parul Verma, Amity University, Lucknow campus, India
Dr. Rachid Alaoui, Agadir Ibn Zohr University, Agadir, Morocco
Dr. Dharmendra Patel, Charotar University of Science and Technology, India
Dr. Dong Zhang, University of Central Florida, USA
Dr. Kennedy Chinedu Okafor, Federal University of Technology Owerri, Nigeria
Prof. C Ram Kumar, Dr NGP Institute of Technology, India
Dr. Sandeep Gupta, GGS IP University, New Delhi, India
Dr. Shahanawaj Ahamad, University of Ha'il, Ha'il City, Ministry of Higher Education, Kingdom of Saudi Arabia
Dr. Najeed Ahmed Khan, NED University of Engineering & Technology, India
Dr. Sajid Ullah Khan, Universiti Malaysia Sarawak, Malaysia
Dr. Muhammad Asif, National Textile University Faisalabad, Pakistan
Dr. Yu BI, University of Central Florida, Orlando, FL, USA
Dr. Brijendra Kumar Joshi, Research Center, Military College of Telecommunication Engineering, India
Prof. Dr. Nak Eun Cho, Pukyong National University, Korea
Prof. Wasim Ul-Haq, Faculty of Science, Majmaah University, Saudi Arabia
Dr. Mohsan Raza, G.C University Faisalabad, Pakistan
Dr. Syed Zakar Hussain Bukhari, National Science and Technology Azad Jamu Kashmir, Pakistan
Dr. Ruksar Fatima, KBN College of Engineering, Gulbarga, Karnataka, India
Associate Professor S. Karpagavalli, Department of Computer Science, PSGR Krishnammal College for Women
Coimbatore, Tamilnadu, India
Dr. Bushra Mohamed Elamin Elhaim, Prince Sattam bin Abdulaziz University, Saudi Arabia
Dr. Shamik Tiwari, Department of CSE, CET, Mody University, Lakshmangarh
Dr. Rohit Raja, Faculty of Engineering and Technology, Shri Shankaracharya Group of Institutions, India
Prof. Dr. Aqeel-ur-Rehman, Department of Computing, HIET, FEST, Hamdard University, Pakistan
Dr. Nageswara Rao Moparthi, Velagapudi Ramakrishna Siddhartha Engineering College, India
Dr. Mohd Muqeem, Department of Computer Application, Integral University, Lucknow, India
Dr. Zeeshan Bhatti, Institute of Information and Communication Technology, University of Sindh, Jamshoro, Pakistan

Dr. Emrah Irmak, Biomedical Engineering Department, Karabuk University, Turkey

Dr. Fouad Abdulameer salman, School of Informatics and Applied Mathematics, Universiti Malaysia Terengganu

Dr. N. Prasath, Department of Computer Science and Engineering, KPR Institute of Engineering and Technology, Arasur, Coimbatore

Dr. Hasan Ashrafi-rizi, Health Information Technology Research Center, Isfahan University of Medical Sciences, Hezar Jerib Avenue, Isfahan, Iran

Dr. N. Sasikaladevi, School of Computing, SASTRA University, Thirumalisamudram, Tamilnadu, India.

Dr. Anchit Bijalwan, Arba Minch University, Ethiopia

Dr. K. Sathishkumar, BlueCrest University College, Accra North, Ghana, West Africa

Dr. Dr. Parameshachari B D, GSSS Institute of Engineering and Technology for Women, Affiliated to Visvesvaraya Technological University, Belagavi

Dr. C. Shoba Bindu, Dept. of CSE, JNTUA College of Engineering, India

Dr. M. Inbavalli, ER. Perumal Manimekalai College of Engineering, Hosur, Tamilnadu, India

Dr. Vidya Sagar Ponnamm, Dept. of IT, Velagapudi Ramakrishna Siddhartha Engineering College, India

Dr. Kelvin LO M. F., The Hong Kong Polytechnic University, Hong Kong

Prof. Karimella Vikram, G.H. Raisoni College of Engineering & Management, Pune, India

Dr. Shajilin Loret J.B., VV College of Engineering, India

Dr. P. Sujatha, Department of Computer Science at Vels University, Chennai

Dr. Vaibhav Sundriyal, Old Dominion University Research Foundation, USA

Dr. Md Masud Rana, Khulna University of Engineering and Technology, Bangladesh

Dr. Gurcharan Singh, Khalsa College Amritsar, Guru Nanak Dev University, Amritsar, India

Dr. Richard Otieno Omollo, Department of Computer Science and Software Engineering, Jaramogi Oginga Odinga University of Science and Technology, Kenya

Prof. (Dr) Amit Verma, Computer Science & Engineering, Chandigarh Engineering College, Landran, Mohali, India

Dr. Vidya Sagar Ponnamm, Velagapudi Ramakrishna Siddhartha Engineering College, India

Dr. Bohui Wang, School of Aerospace Science and Technology, Xidian University, P.R. China

Dr. M. Anjan Kumar, Department of Computer Science, Satavahana University, Karimnagar

Dr. Hanumanthappa J., DoS in CS, Uni of Mysuru, Karnataka, India

Dr. Pouya Derakhshan-Barjoei, Dept. of Telecommunication and Engineering, Islamic Azad University, Iran

Professor Edelberto Silva, Universidade Federal de Juiz de Fora, Brazil

Dr. Sonali Vyas, Amity University Rajasthan, India

Dr. Santosh Bharti, National Institute of Technology Rourkela, India

Dr. Deepak Gupta, Maharaja Agrasen Institute of Technology, India

Dr. Emrah Irmak, Karabuk University, Turkey

Dr. Yojna Arora, Amity University, India

Dr. Marta Cimitile, Unitelma Sapienza, Italy

Assistant Prof. Shanthakumari Raju, Kongu Engineering College, India

Dr. Ravi Verma, RGPV Bhopal, India

Dr. Tanweer Alam, Islamic University of Madinah, Dept. of Computer Science, College of Computer and Information System, Al Madinah, Saudi Arabia

Dr. Kumar Keshamoni, Dept. of ECE, Vaagdevi Engineering College, Warangal, Telangana, India

Dr. G. Rajkumar, N.M.S.S.Vellaichamy Nadar College, Madurai, Tamilnadu, India

Dr. P. Mayil Vel Kumar, Karpagam Institute of Technology, Coimbatore, India

Dr. M. Yaswanth Bhanu Murthy, Vasireddy Venkatadri Institute of Technology, Guntur, A.P., India

Asst. Prof. Dr. Mehmet Barış TABAKCIOĞLU, Bursa Technical University, Turkey

Dr. Mohd. Muntjir, College of Computers and Information Technology, Taif University, Kingdom of Saudi Arabia

Dr. Sanjay Agal, Aravali Institute of Technical Studies, Udaipur, India

Dr. Shanshan Tuo, xAd Inc., US
Dr. Subhadra Shaw, AKS University, Satna, India
Dr. Piyush Anand, Noida International University, Greater Noida, India
Dr. Brijendra Kumar Joshi, Research Center Military College of Telecommunication Engineering, India
Dr. V. Sreerama Murthy, GMRIT, Rajam, AP, India
Dr. S. Nagarajan, Annamalai University, India
Prof. Pramod Bhausaheb Deshmukh, D. Y. Patil College of Engineering, Akurdi, Pune, India
Dr. Jaspreet Kour, GCET, India
Dr. Parul Agarwal, Jamia Hamdard
Dr. Muhammad Faheem, Abduallah Gul University
Dr. Vaibhav Sundriyal, Old Dominion University
Dr. Sujatha Dandu, JNTUH
Dr. Wenzhao Zhang, NCSU, US
Dr. Senthil Kumar P., Anna University
Dr. Harshal Karande, Arvind Gavali College of Engineering, Satara
Dr. Kannan Dhandapani, Nehru Arts and Science College, Affiliated to Bharatiar Univerisity
Prof. Dr. Muthukumar Subramnian, Indian Institute of Information Technology, Tamilnadu, India
Dr. K .Vengatesan Krishnasamy, Dr. BATU University
Dr. Jayapandian N., Knowledge Institute of Technology
Dr. Sangeetha S.K.B, Rajalakshmi Engineering College
Dr. Geetha Devi Appari, PVP Siddhartha Institute of Technology
Dr. Pradeep Gurunathan, A.V.C. College of Engineering
Dr. Muftah Fraifer, Interaction design Center-University of Limerick
Dr. Gamal Eladl, Mansoura University/ IS Dept.
Dr. Bereket Assa, Woliyta Soddo University
Dr. Venkata Suryanarayana Tinnaluri, Malla Reddy Group of Institutions
Dr. Jagadeesh Gopal, VIT University, Vellore
Dr. Vidya Sagar Ponnamm, JNTUK, Kakinada/Velagapudi Ramakrishna Siddhartha Engineering College
Dr. Meenashi Sharma, Chandigarh University
Dr. Hiyam Hatem, University of Baghdad, College of Science
Dr. Smitha Elsa Peter, PRIST University
Dr. Gurcharan Singh, Guru Nanak Dev University
Dr. Ahmed EL-YAHYAOU, Mohammed V University in Rabat
Dr. Shruti Bahrgava, JNTUH
Dr. Seda Kul, Kocaeli University
Dr. Bappaditya Jana, Chaibasa Engineering College
Dr. Farhad Goodarzi, UPM university
Dr. Sujatha P., Vels University, Chennai
Dr. Satya Bhushan Verma, National Institute of Technology Durgapur
Dr. Man Fung LO, The Hong Kong Polytechnic University
Dr. Muhammad Adnan, Abdul Wali Khan University
Dr. Seyed Sahand Mohammadi Ziabari, Vrije University
Dr. Brindha Srinivasan, Palanisamy College of Arts, Erode
Dr. Mohammad Aldabbagh, University of Mosul
Prof. Abdallah Rhattoy, Moulay Ismail University, Higher School of Technology
Dr. Kumar Keshamoni, Vaagdevi Engineering College, Warangal, Telangana, India
Dr. Khalid Nazim Abdus Sattar, College of Science, Az-Zulfi campus, Majmaah university, Kingdom of Saudi Arabia

CALL FOR PAPERS

International Journal of Computer Science and Information Security

IJCSIS 2018-2019

ISSN: 1947-5500

<http://sites.google.com/site/ijcsis/>

International Journal Computer Science and Information Security, IJCSIS, is the premier scholarly venue in the areas of computer science and security issues. IJCSIS 2011 will provide a high profile, leading edge platform for researchers and engineers alike to publish state-of-the-art research in the respective fields of information technology and communication security. The journal will feature a diverse mixture of publication articles including core and applied computer science related topics.

Authors are solicited to contribute to the special issue by submitting articles that illustrate research results, projects, surveying works and industrial experiences that describe significant advances in the following areas, but are not limited to. Submissions may span a broad range of topics, e.g.:

Track A: Security

Access control, Anonymity, Audit and audit reduction & Authentication and authorization, Applied cryptography, Cryptanalysis, Digital Signatures, Biometric security, Boundary control devices, Certification and accreditation, Cross-layer design for security, Security & Network Management, Data and system integrity, Database security, Defensive information warfare, Denial of service protection, Intrusion Detection, Anti-malware, Distributed systems security, Electronic commerce, E-mail security, Spam, Phishing, E-mail fraud, Virus, worms, Trojan Protection, Grid security, Information hiding and watermarking & Information survivability, Insider threat protection, Integrity

Intellectual property protection, Internet/Intranet Security, Key management and key recovery, Language-based security, Mobile and wireless security, Mobile, Ad Hoc and Sensor Network Security, Monitoring and surveillance, Multimedia security, Operating system security, Peer-to-peer security, Performance Evaluations of Protocols & Security Application, Privacy and data protection, Product evaluation criteria and compliance, Risk evaluation and security certification, Risk/vulnerability assessment, Security & Network Management, Security Models & protocols, Security threats & countermeasures (DDoS, MiM, Session Hijacking, Replay attack etc.), Trusted computing, Ubiquitous Computing Security, Virtualization security, VoIP security, Web 2.0 security, Submission Procedures, Active Defense Systems, Adaptive Defense Systems, Benchmark, Analysis and Evaluation of Security Systems, Distributed Access Control and Trust Management, Distributed Attack Systems and Mechanisms, Distributed Intrusion Detection/Prevention Systems, Denial-of-Service Attacks and Countermeasures, High Performance Security Systems, Identity Management and Authentication, Implementation, Deployment and Management of Security Systems, Intelligent Defense Systems, Internet and Network Forensics, Large-scale Attacks and Defense, RFID Security and Privacy, Security Architectures in Distributed Network Systems, Security for Critical Infrastructures, Security for P2P systems and Grid Systems, Security in E-Commerce, Security and Privacy in Wireless Networks, Secure Mobile Agents and Mobile Code, Security Protocols, Security Simulation and Tools, Security Theory and Tools, Standards and Assurance Methods, Trusted Computing, Viruses, Worms, and Other Malicious Code, World Wide Web Security, Novel and emerging secure architecture, Study of attack strategies, attack modeling, Case studies and analysis of actual attacks, Continuity of Operations during an attack, Key management, Trust management, Intrusion detection techniques, Intrusion response, alarm management, and correlation analysis, Study of tradeoffs between security and system performance, Intrusion tolerance systems, Secure protocols, Security in wireless networks (e.g. mesh networks, sensor networks, etc.), Cryptography and Secure Communications, Computer Forensics, Recovery and Healing, Security Visualization, Formal Methods in Security, Principles for Designing a Secure Computing System, Autonomic Security, Internet Security, Security in Health Care Systems, Security Solutions Using Reconfigurable Computing, Adaptive and Intelligent Defense Systems, Authentication and Access control, Denial of service attacks and countermeasures, Identity, Route and

Location Anonymity schemes, Intrusion detection and prevention techniques, Cryptography, encryption algorithms and Key management schemes, Secure routing schemes, Secure neighbor discovery and localization, Trust establishment and maintenance, Confidentiality and data integrity, Security architectures, deployments and solutions, Emerging threats to cloud-based services, Security model for new services, Cloud-aware web service security, Information hiding in Cloud Computing, Securing distributed data storage in cloud, Security, privacy and trust in mobile computing systems and applications, **Middleware security & Security features:** middleware software is an asset on

its own and has to be protected, interaction between security-specific and other middleware features, e.g., context-awareness, **Middleware-level security monitoring and measurement:** metrics and mechanisms for quantification and evaluation of security enforced by the middleware, **Security co-design:** trade-off and co-design between application-based and middleware-based security, **Policy-based management:** innovative support for policy-based definition and enforcement of security concerns, **Identification and authentication mechanisms:** Means to capture application specific constraints in defining and enforcing access control rules, **Middleware-oriented security patterns:** identification of patterns for sound, reusable security, **Security in aspect-based middleware:** mechanisms for isolating and enforcing security aspects, **Security in agent-based platforms:** protection for mobile code and platforms, Smart Devices: Biometrics, National ID cards, Embedded Systems Security and TPMs, RFID Systems Security, Smart Card Security, Pervasive Systems: Digital Rights Management (DRM) in pervasive environments, Intrusion Detection and Information Filtering, Localization Systems Security (Tracking of People and Goods), Mobile Commerce Security, Privacy Enhancing Technologies, Security Protocols (for Identification and Authentication, Confidentiality and Privacy, and Integrity), Ubiquitous Networks: Ad Hoc Networks Security, Delay-Tolerant Network Security, Domestic Network Security, Peer-to-Peer Networks Security, Security Issues in Mobile and Ubiquitous Networks, Security of GSM/GPRS/UMTS Systems, Sensor Networks Security, Vehicular Network Security, Wireless Communication Security: Bluetooth, NFC, WiFi, WiMAX, WiMedia, others

This Track will emphasize the design, implementation, management and applications of computer communications, networks and services. Topics of mostly theoretical nature are also welcome, provided there is clear practical potential in applying the results of such work.

Track B: Computer Science

Broadband wireless technologies: LTE, WiMAX, WiRAN, HSDPA, HSUPA, Resource allocation and interference management, Quality of service and scheduling methods, Capacity planning and dimensioning, Cross-layer design and Physical layer based issue, Interworking architecture and interoperability, Relay assisted and cooperative communications, Location and provisioning and mobility management, Call admission and flow/congestion control, Performance optimization, Channel capacity modeling and analysis, Middleware Issues: Event-based, publish/subscribe, and message-oriented middleware, Reconfigurable, adaptable, and reflective middleware approaches, Middleware solutions for reliability, fault tolerance, and quality-of-service, Scalability of middleware, Context-aware middleware, Autonomic and self-managing middleware, Evaluation techniques for middleware solutions, Formal methods and tools for designing, verifying, and evaluating, middleware, Software engineering techniques for middleware, Service oriented middleware, Agent-based middleware, Security middleware, Network Applications: Network-based automation, Cloud applications, Ubiquitous and pervasive applications, Collaborative applications, RFID and sensor network applications, Mobile applications, Smart home applications, Infrastructure monitoring and control applications, Remote health monitoring, GPS and location-based applications, Networked vehicles applications, Alert applications, Embedded Computer System, Advanced Control Systems, and Intelligent Control : Advanced control and measurement, computer and microprocessor-based control, signal processing, estimation and identification techniques, application specific IC's, nonlinear and adaptive control, optimal and robot control, intelligent control, evolutionary computing, and intelligent systems, instrumentation subject to critical conditions, automotive, marine and aero-space control and all other control applications, Intelligent Control System, Wiring/Wireless Sensor, Signal Control System. Sensors, Actuators and Systems Integration : Intelligent sensors and actuators, multisensor fusion, sensor array and multi-channel processing, micro/nano technology, microsensors and microactuators, instrumentation electronics, MEMS and system integration, wireless sensor, Network Sensor, Hybrid

Sensor, Distributed Sensor Networks. Signal and Image Processing : Digital signal processing theory, methods, DSP implementation, speech processing, image and multidimensional signal processing, Image analysis and processing, Image and Multimedia applications, Real-time multimedia signal processing, Computer vision, Emerging signal processing areas, Remote Sensing, Signal processing in education. Industrial Informatics: Industrial applications of neural networks, fuzzy algorithms, Neuro-Fuzzy application, bioInformatics, real-time computer control, real-time information systems, human-machine interfaces, CAD/CAM/CAT/CIM, virtual reality, industrial communications, flexible manufacturing systems, industrial automated process, Data Storage Management, Harddisk control, Supply Chain Management, Logistics applications, Power plant automation, Drives automation. Information Technology, Management of Information System : Management information systems, Information Management, Nursing information management, Information System, Information Technology and their application, Data retrieval, Data Base Management, Decision analysis methods, Information processing, Operations research, E-Business, E-Commerce, E-Government, Computer Business, Security and risk management, Medical imaging, Biotechnology, Bio-Medicine, Computer-based information systems in health care, Changing Access to Patient Information, Healthcare Management Information Technology. Communication/Computer Network, Transportation Application : On-board diagnostics, Active safety systems, Communication systems, Wireless technology, Communication application, Navigation and Guidance, Vision-based applications, Speech interface, Sensor fusion, Networking theory and technologies, Transportation information, Autonomous vehicle, Vehicle application of affective computing, Advance Computing technology and their application : Broadband and intelligent networks, Data Mining, Data fusion, Computational intelligence, Information and data security, Information indexing and retrieval, Information processing, Information systems and applications, Internet applications and performances, Knowledge based systems, Knowledge management, Software Engineering, Decision making, Mobile networks and services, Network management and services, Neural Network, Fuzzy logics, Neuro-Fuzzy, Expert approaches, Innovation Technology and Management : Innovation and product development, Emerging advances in business and its applications, Creativity in Internet management and retailing, B2B and B2C management, Electronic transceiver device for Retail Marketing Industries, Facilities planning and management, Innovative pervasive computing applications, Programming paradigms for pervasive systems, Software evolution and maintenance in pervasive systems, Middleware services and agent technologies, Adaptive, autonomic and context-aware computing, Mobile/Wireless computing systems and services in pervasive computing, Energy-efficient and green pervasive computing, Communication architectures for pervasive computing, Ad hoc networks for pervasive communications, Pervasive opportunistic communications and applications, Enabling technologies for pervasive systems (e.g., wireless BAN, PAN), Positioning and tracking technologies, Sensors and RFID in pervasive systems, Multimodal sensing and context for pervasive applications, Pervasive sensing, perception and semantic interpretation, Smart devices and intelligent environments, Trust, security and privacy issues in pervasive systems, User interfaces and interaction models, Virtual immersive communications, Wearable computers, Standards and interfaces for pervasive computing environments, Social and economic models for pervasive systems, Active and Programmable Networks, Ad Hoc & Sensor Network, Congestion and/or Flow Control, Content Distribution, Grid Networking, High-speed Network Architectures, Internet Services and Applications, Optical Networks, Mobile and Wireless Networks, Network Modeling and Simulation, Multicast, Multimedia Communications, Network Control and Management, Network Protocols, Network Performance, Network Measurement, Peer to Peer and Overlay Networks, Quality of Service and Quality of Experience, Ubiquitous Networks, Crosscutting Themes – Internet Technologies, Infrastructure, Services and Applications; Open Source Tools, Open Models and Architectures; Security, Privacy and Trust; Navigation Systems, Location Based Services; Social Networks and Online Communities; ICT Convergence, Digital Economy and Digital Divide, Neural Networks, Pattern Recognition, Computer Vision, Advanced Computing Architectures and New Programming Models, Visualization and Virtual Reality as Applied to Computational Science, Computer Architecture and Embedded Systems, Technology in Education, Theoretical Computer Science, Computing Ethics, Computing Practices & Applications

Authors are invited to submit papers through e-mail ijcsiseditor@gmail.com. Submissions must be original and should not have been published previously or be under consideration for publication while being evaluated by IJCSIS. Before submission authors should carefully read over the journal's Author Guidelines, which are located at <http://sites.google.com/site/ijcsis/authors-notes> .



© IJCSIS PUBLICATION 2018

ISSN 1947 5500

<http://sites.google.com/site/ijcsis/>